

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/135018>

Copyright and reuse:

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

**Cost-effectiveness of innovations in pathology services in relation to
cancer diagnosis and treatment management**

Asmaa El-Banna

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

University of Warwick

June 2019

Table of contents

List of figures	vi
List of tables.....	viii
Acknowledgements.....	ix
Declaration	x
Dedication.....	xi
Abstract	xii
List of Abbreviations	xiii
Chapter 1: Introduction.....	1
1.1 Cancer	1
1.1.1 Cancer outcomes.....	1
1.1.2 Cancer pathways	2
1.1.3 Government policy.....	5
1.2 Pathology.....	8
1.2.1 Challenges in pathology	9
1.2.2 Managing demand	10
1.2.3 Pathology innovation	10
1.3 Economic evaluations	12
1.3.1 Challenges in the evaluation of pathology services.....	13
Chapter 2: Economic evaluations of pathology innovations - a systematic review	15
2.1 Introduction.....	15
2.1.1 Economic evaluations within pathology	15
2.2 Aims	16
2.3 Methods.....	17
2.3.1 Information sources	17
2.3.2 Eligibility criteria.....	18
2.3.3 Data collection and analysis	20
2.3.4 Data synthesis	21
2.4 Results	22
2.4.1 Search results	22
2.4.2 General study characteristics.....	23
2.4.3 Economic evaluation methods adopted.....	25
2.4.4 Cost-effectiveness results reported.....	31
2.4.5 Quality assessment	33
2.5 Discussion and conclusion	34
Chapter 3: An introduction to digital pathology	37

3.1 Speed	38
3.1.1 The laboratory workflow	39
3.1.2 MDT meetings	43
3.1.3 Reporting beyond the laboratory	45
3.1.4 The patient experience	46
3.2 Accuracy.....	47
3.2.1 Biomarker profile	50
3.2.2 Lymph node status	52
3.2.3 Tumour grade	53
3.2.4 Tumour size	55
3.3 Conclusion	55
Chapter 4: The breast cancer pathway.....	57
4.1 Introduction.....	57
4.1.1 Information sources	57
4.2 Aims	59
4.3 Methods.....	60
4.3.1 Methods for the descriptions of the evidence-based pathway	60
4.3.2 Methods for the description of the clinical-practice pathway	62
4.3.3 Methods for the description of the network of digital pathology processes	63
4.4 Results	64
4.4.1 Evidence-based pathway	64
4.4.2 Clinical practice pathway	74
4.4.3 Relationship between digital pathology and breast cancer outcomes	82
4.5 Discussion and conclusion	84
Chapter 5: Economic evaluation of digital pathology.....	87
5.1 Introduction.....	87
5.1.1 Decision modelling for economic evaluation	88
5.2 Aims	90
5.3 Methods.....	91
5.3.1 DES model development.....	91
5.3.2 Economic evaluation	118
5.4 Results	127
5.4.1 Model validation	127
5.4.2 Conventional microscope outputs.....	128
5.4.3 Digital pathology outputs	131
5.5 Discussion	134

5.6 Conclusion	137
Chapter 6: Systematic review and meta-analysis of HER2 testing accuracy.....	138
6.1 Introduction.....	138
6.1.1 HER2 amplification and overexpression.....	138
6.1.2 Anti-HER2 therapy.....	139
6.1.3 Approaches to HER2 testing	139
6.1.4 HER2 testing algorithm	141
6.1.5 Testing thresholds	142
6.1.6 Accuracy of HER2 tests.....	143
6.1.7 Quantifying accuracy.....	144
6.2 Aims	147
6.3 Methods.....	148
6.3.1 Information sources	148
6.3.2 Eligibility criteria.....	149
6.3.3 Data collection and analysis	151
6.3.4 Data synthesis	152
6.4 Results	154
6.4.1 Study characteristics	155
6.4.2 Quality assessment	159
6.4.3 Accuracy data extraction	161
6.4.4 Base case analysis	166
6.4.5 Sensitivity analysis	168
6.5 Discussion	174
6.6 Conclusion	179
Chapter 7: Economic evaluation of digital HER2 testing.....	180
7.1 Introduction.....	180
7.1.1 Patient care	180
7.1.2 Economic evaluation	180
7.2 Aims	181
7.3 Methods.....	182
7.3.1 Model structure.....	182
7.3.2 Model costs	184
7.3.3 Model QALYs	185
7.3.4 UK prevalence	186
7.3.5 Model probabilities	186
7.3.6 Base case scenario.....	187

7.3.7 Scenario analysis 1: 2008 scoring thresholds	189
7.3.8 Scenario analysis 2: 12 month treatment period.....	189
7.3.9 Scenario analysis 3: Single study approach	189
7.3.10 Cost-effectiveness analysis	190
7.3.11 Value of Information Analysis	191
7.4 Results	192
7.4.1 Base case analysis results	192
7.4.2 Scenario analysis results	194
7.4.3 Value of Information Analysis	196
7.5 Discussion	199
7.6 Conclusion	202
Chapter 8: General discussion and conclusion.....	203
8.1 Overview of thesis and main findings	203
8.1.1 The breast cancer pathway.....	203
8.1.2 Economic evaluation using DES modelling.....	204
8.1.3 HER2 IHC accuracy with digital pathology.....	207
8.1.4 Economic evaluation of digital pathology for HER2 scoring.....	208
8.2 Strengths and limitations of research	209
8.3 Recommendations for policy	209
8.4 Recommendations for research.....	210
8.5 Conclusions	210
References	211
Appendices	231
Appendix A: Search strategies-systematic review of economic evaluations.....	231
Appendix B: Data extraction form - systematic review of economic evaluations.....	235
Appendix C: Simul8 building blocks.....	236
Appendix D: List of search terms	237
Appendix E: Search terms used in initial phase search strategy	238
Appendix F: Search strategies - systematic review of HER2 accuracy.....	239
Appendix G: Data extraction form - systematic review of HER2 accuracy	242

List of figures

Figure 1-1 General cancer pathway	2
Figure 1-2 Cancer waiting time targets.....	7
Figure 1-3 Relationship between pathology processes and patient outcomes.....	14
Figure 2-1 Study selection process	22
Figure 2-2 Summary of cancer types investigated.....	24
Figure 2-3 Summary of papers by country of origin	24
Figure 2-4 Summary pathology interventions investigated.....	25
Figure 2-5 Summary modelling approaches adopted.....	26
Figure 2-6 Distribution of CHEERS scores for all of the eligible studies.....	33
Figure 3-1 Summary of the speed impacts of digital pathology	38
Figure 3-2 Conventional microscope versus digital pathology processes	40
Figure 3-3 Overview of the patient cancer pathway	46
Figure 3-4 Summary of the areas of potential accuracy gains with DP	49
Figure 4-1 Steps in suspected breast cancer management based on published guidelines .	65
Figure 4-2 Guidelines for triple assessment	66
Figure 4-3 Overview of surgery and radiotherapy guidelines for DCIS cases.....	67
Figure 4-4 Summary of the recommended surgical pathway for IBC cases	69
Figure 4-5 Overview of the radiotherapy guidelines for IBC cases.....	70
Figure 4-6 ER and HER2 therapy treatment guidelines	71
Figure 4-7 Overview of the clinical practice breast cancer pathway.....	75
Figure 4-8 Outpatient clinic (symptomatic patients).....	76
Figure 4-9 Breast screening unit (symptomatic patients).....	76
Figure 4-10 Breast screening unit (asymptomatic patients).....	77
Figure 4-11 Diagnostic MDT decision algorithm.....	79
Figure 4-12 Neoadjuvant therapy.....	80
Figure 4-13 Adjuvant therapy	81
Figure 4-14 Post-surgery MDT	82
Figure 4-15 Overview of process chain network for digital pathology.....	83
Figure 5-1 Steps in DES model development.....	92
Figure 5-2 Summary breast cancer pathway through secondary care	94
Figure 5-3 Overview of the conceptual model	97
Figure 5-4 DES model structure in Simul8	106
Figure 5-5 Frequency of breast core arrivals at UHCW	113
Figure 5-6 Experimental factors.....	120
Figure 5-7 Days from CB arrival in pathology to adjuvant treatment plan.....	127
Figure 5-8 Two sample t-test comparing data and model output.....	127
Figure 5-9 Histogram of the number of times an invasive breast cancer case is discussed at the MDT meeting	128
Figure 5-10 Histogram of the number of MDT discussions/IBC patient.....	132
Figure 5-11 Mean number of MDT discussions/IBC patient.....	132
Figure 6-1 HER2 receptor expression in normal and cancerous cells.....	139
Figure 6-2 HER2 testing algorithm	141
Figure 6-3 Study identification and selection process.....	154
Figure 6-4 Sensitivity and specificity of HER2 IHC using CM.....	164
Figure 6-5 Sensitivity and specificity of HER2 IHC using DP IA.....	165
Figure 6-6 BCA comparison of CM and DP IA with SROC curves	166

Figure 6-7 BCA proportions of equivocal scores if HER2 positive: CM	167
Figure 6-8 BCA proportion of equivocal scores if HER2 positive: DP IA.....	167
Figure 6-9 BCA Proportion of equivocal scores if HER2 negative: CM.....	168
Figure 6-10 BCA proportion of equivocal scores if HER2 negative: DP IA	168
Figure 6-11 SA1 Comparison of CM and DP IA with SROC curves	169
Figure 6-12 SA1 Proportion of equivocal cases if HER2 positive: CM.....	169
Figure 6-13 SA1 Proportion of equivocal cases if HER2 positive: DP IA.....	170
Figure 6-14 SA1 Proportion of equivocal cases if HER2 negative: CM.....	170
Figure 6-15 SA1 Proportion of equivocal cases if HER2 negative: DP IA	170
Figure 6-16 SA2 Comparison of CM and DP IA with SROC curves	171
Figure 6-17 SA2 Proportion of equivocal cases if HER2 positive: CM.....	171
Figure 6-18 SA2 Proportion of equivocal cases if HER2 positive: DP IA.....	172
Figure 6-19 SA2 Proportion of equivocal cases if HER2 negative: CM.....	172
Figure 6-20 SA2 Proportion of equivocal cases if HER2 negative: DP IA	172
Figure 6-21 Comparison of DP IA technologies with SROC curves	173
Figure 7-1 Illustration of the convention microscope arm of the decision tree.....	183
Figure 7-2 BCA cost-effectiveness plane.....	193
Figure 7-3 BCA CEAC	193
Figure 7-4 SA1 cost-effectiveness plane	194
Figure 7-5 SA2 cost-effectiveness plane	194
Figure 7-6 SA3 Cost-effectiveness plane.....	195
Figure 7-7 CEAC for all scenarios	195
Figure 7-8 Patient level EVPI	197

List of tables

Table 2-1 Summary inclusion criteria	19
Table 2-2 HER2 testing strategies compared by Dendukuri et al. (2007)	29
Table 3-1 Agreement between DP and CM when grading breast cancers	54
Table 4-1 List of publications to inform the evidence-based breast cancer pathway.....	61
Table 4-2 Summary of indications for chemotherapy in breast cancer	73
Table 5-1 Summary of simplifications in DES model	95
Table 5-2 Summary of pathology CB results.....	100
Table 5-3 Management of B3/B4 core biopsies in the dataset	102
Table 5-4 Number of breast tissue specimens per patient.....	108
Table 5-5 UHCW data collected	109
Table 5-6 Model probabilities	110
Table 5-7 Model time parameters in days.....	112
Table 5-8 Model cost parameters.....	114
Table 5-9 Summary of chemotherapy regimens at UHCW.....	117
Table 5-10 Summary impacts of Digital Pathology	120
Table 5-11 Accuracy of breast cancer grade reporting.....	122
Table 5-12 Experimental scenarios	123
Table 5-13 Patient groups to be analysed independently	124
Table 5-14 Simul8 trial calculator results.....	125
Table 5-15 Model outputs under each scenario.....	129
Table 6-1 HER2 scoring thresholds, 2007 versus 2013	143
Table 6-2 Example 2x2 table	145
Table 6-3 Systematic review of accuracy - summary inclusion criteria	151
Table 6-4 Summary of included studies.....	157
Table 6-5 Summary digital pathology image analysis methods adopted in each study.....	158
Table 6-6 Quality assessment of included studies.....	160
Table 6-7 Summary change in point estimates of sensitivity (SN), specificity (SP) and proportion of equivocal results reported in each study when comparing DP to CM.....	163
Table 6-8 Sensitivity and specificity estimates by DP IA technology	173
Table 7-1 Decision tree endpoints	183
Table 7-2 Cost of Herceptin treatment (Clarke et al., 2017)	185
Table 7-3 Example 9 week mean model costs input.....	185
Table 7-4 QALYs of Herceptin treatment.....	186
Table 7-5 List of test effects	187
Table 7-6 Estimates of Log(OR)	187
Table 7-7 BCA p1 distributions.....	188
Table 7-8 SA1 p1 distributions	189
Table 7-9 SA3 p1 distributions	190
Table 7-10 Deterministic model outputs	192
Table 7-11 Probability of cost-effectiveness at NICE WTP thresholds	196
Table 7-12 EVPI at NICE WTP thresholds	197
Table 7-13 Population level EVPI	198
Table 7-14 EVPPI results	198

Acknowledgements

First and foremost all praise and gratitude to God, the most gracious the most merciful, for providing me with the ability and the strength to complete this research. Without His constant presence, this research would not have been possible.

I am deeply grateful to my supervisor Professor Jason Madan, who has supported and encouraged me extensively throughout this PhD journey in addition to continually sharing his invaluable expertise and knowledge. I would also like to thank my other supervisors, Professor Ian Cree and Professor Stavros Petrou.

I am indebted to my colleagues at the University of Warwick. I am grateful to Peter Auguste who offered his insight into the options for meta-analysis of sensitivity and specificity. I would also like to thank Dr Felix Achana for his help and suggestions around the use of test performance measures in economic evaluations.

This thesis would not have been possible without the support of my family and friends, who believed in me when I no longer believed in myself. In particular, I would like to express my gratitude to my parents, Dr Hany El-Banna and Mrs. Youseria Labib for their love and continued support and for always being the best of role models, I am forever indebted to them. To my siblings, Fatima, Hassan, Maryam and Omar for their continuous encouragement and to Ayman and Amara, for always making me smile. I am forever grateful to my friends Samantha and Bishal, we started this journey together six years ago and I hope our friendship lasts for many more years to come.

Finally yet importantly, this research would not have been possible without the understanding and support of my husband Dr Usama Attia. Thank you for always believing in me.

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

Dedication

To my husband Usama,
for your unwavering love, patience and support throughout this PhD journey

Abstract

Pathology plays an important role in cancer diagnosis and treatment management, results from the pathology lab guide clinicians' diagnosis and inform patient care plans. Pathology digitisation is expected to maximise lab efficiency when handling tissue specimens, enhance speed, provide novel information to be used by clinicians when making treatment decisions and potentially improve test accuracy. Early cancer diagnosis and personalised treatment are key players in enhancing patients' clinical outcomes and improving quality of life. Whilst research has shown digitisation of pathology labs to be an effective intervention for better management and reporting on tissue specimens, no evaluation has reported on the economic implications of the adoption of digital systems in an NHS with limited resources. Breast cancers are the most common cancer type in the UK so any advances in accuracy or time to diagnosis due to digital pathology are expected to have a large impact on this group of patients.

This thesis investigates the cost-effectiveness of digital pathology through its impacts on breast cancer patients. A discrete event simulation model representing the breast cancer pathway was constructed and used to analyse the impacts of digitisation. There was evidence of both time and cost savings for breast cancer patients as a result of pathology digitisation. A systematic review and meta-analysis compared the diagnostic accuracy of the HER2 biomarker pre- and post- the introduction of digital pathology. There was evidence of reporting precision but not of improved accuracy. Finally, a cost-effectiveness analysis comparing the two approaches found digital pathology not to be cost-effective when compared to conventional microscopes for scoring the HER2 biomarker.

List of Abbreviations

ALND	Axillary lymph node dissection
ASCO/CAP	American Society of Clinical Oncology/College of American Pathologists
AUC	Area under receiver operating curve
BCA	Base case analysis
CAMELYON16	Cancer metastases in lymph node challenge 2016
CB	Core biopsy
CBA	Cost-benefit analysis
CCA	Cost-consequence analysis
CEA	Cost-effectiveness analysis
CEAC	Cost-effectiveness acceptability curve
CHEERS	Consolidated Health Economic Evaluation Reporting Standards
CM	Conventional microscope
CMA	Cost-minimisation analysis
CI	Confidence interval
CRD	Centre for Reviews and Dissemination
CUA	Cost-utility analysis
CyT	Cytotechnologist
DCIS	Ductal carcinoma in situ
DEB	Diagnostic excision biopsy
DES	Discrete event simulation
DP	Digital pathology
EQ	Equivocal
ER	Oestrogen receptor
EVPI	Expected value of perfect information
EVPPi	Expected value of partial perfect information
FISH	Fluorescence in situ hybridisation
FN	False negative
FP	False positive
FPR	false positive rate
GP	General Practitioner

HER2	Human epidermal growth factor receptor 2
HERA	HERceptin Adjuvant
HSROC	Hierarchical summary receiver operating characteristic
IA	Image analysis
IBC	Invasive breast cancer
ICER	Incremental cost-effectiveness ratio
IHC	Immunohistochemistry
INB	Incremental Net Benefit
Inc	Incremental
IPE	Intra-operative pathology examination
LN	Lymph node
LNB	lymph node biopsy
MDT	Multi-disciplinary team
NEQ	Not equivocal
NHSBSP	NHS Breast screening Programme
NICE	National Institute for Health and Care Excellence
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
OPA	Out-patient appointments
OR	Odds ratio
P	Pathologist
PHE	Public Health England
PR	Progesterone receptor
PRISMA	Preferred reporting items for systematic reviews and meta-analyses
QALY	Quality-adjusted life-year
QUADAS	Quality assessment of diagnostic accuracy studies
RCPATH	The Royal College of Pathologists
RCT	Randomised controlled trial
SA	Sensitivity analysis
SE	Standard error
SLN	Sentinel lymph node

SLNB	Sentinel lymph node biopsy
SN	Sensitivity
SP	Specificity
SROC	Summary receiver operating curve
TA	Technology appraisal
TAT	Turnaround time
TMA	Tissue micro-arrays
TN	True negative
TP	True positive
UHCW	University Hospitals Coventry and Warwickshire
WLE	Wide local excision
WM	Whole-mount
WS	Whole slides
WTP	Willingness to pay
VOI	Value of information

Chapter 1: Introduction

1.1 Cancer

Cancer is classified as one of the five main causes of avoidable deaths in England (Department of Health, 2014), it accounted for 28% of UK deaths in 2016 (Cancer Research UK, 2016a), is related to over 25% of deaths in Europe (eurostat, 2018) and is the second leading cause of death worldwide (WHO, 2018). While cancer 5-year survival has increased steadily in England over the 40 years up to 2011, the number of new cases has more than doubled over this period (ONS, 2015), driven mostly by an ageing demographic.

Cancer can come in many forms, subject to the part of the body affected and the cell types within these that show evidence of malignancy. Furthermore, cancer variants exist across an individual organ depending on the specific underlying characteristics of the disease. Some cancers are more common such as breast, lung and prostate cancer whilst others are categorised as rare (Cancer Research UK, 2018). Some cancers are associated with better chances of survival such as cervical and prostate cancer whilst others are associated with greater rates of mortality, for instance pancreatic cancer (Quaresma et al., 2015). The cancer type and the stage at diagnosis in part drive outcomes for cancer patients; however, health service systems and pathways in place to manage suspected cancer cases have an important role to play in maintaining and improving outcomes for cancer patients.

1.1.1 Cancer outcomes

Disparities in cancer outcomes exist between countries, owing to variations in healthcare systems (Starfield et al., 2005), patient's access to tests and investigations of suspected cancer (Rose et al., 2015), patient's awareness of the symptoms and risk of cancer (Forbes et al., 2013) and the primary research levels within countries to support evidence based practice (Are et al., 2018), to name but a few. Whilst perhaps greater differences in these contributory factors and others not listed are expected when comparing outcomes of low and middle-income countries to those of high-income countries, still disparities in cancer survival also exist among high-income countries.

Cancer survival rates in England are inferior to those of similar countries (Foot and Harrison, 2011), when compared to the rest of Europe, 5-year cancer survival for cancers diagnosed between 1991 and 1995 was generally below the European average (Coleman et al., 2003). This trend continued when estimating and comparing 5-year survival across Europe for cancers diagnosed between 1995 and 2007 (De Angelis et al., 2014). Approximately 7000 cancer deaths would have been avoided each year between 1985 and 1999 if survival rates

in the UK were equal to the average European rates (Abdel-Rahman et al., 2009). Beyond Europe, UK survival for cancers diagnosed between the years 2000 and 2014 has repeatedly fallen behind that of Australia, Canada, New Zealand and the United States (Nuffield trust, 2018).

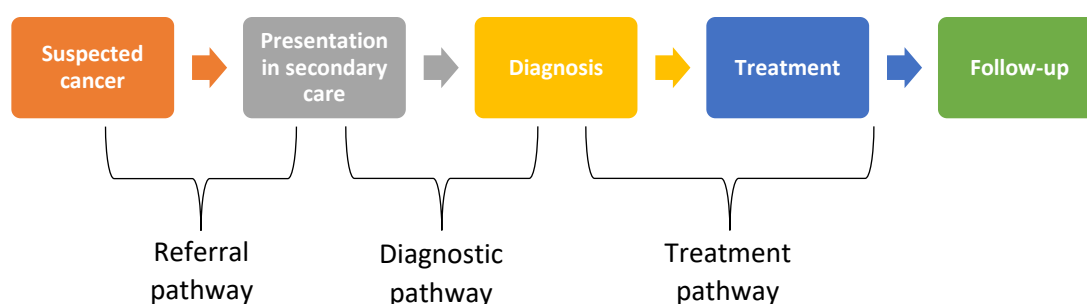
Several factors contribute to the poor survival outcomes reported in the UK, yet published research has continually shown delays in the time to diagnosis as a prominent issue, these can be delays in patient presentation in primary care with symptoms of suspected cancer, delays in referral from primary to secondary care or delays in investigation, diagnosis and treatment in secondary care (Majeed et al., 2018). No doubt an earlier diagnosis of cancer would result in fewer preventable deaths (Hiom, 2015). In addition to time delays, differences in the treatment received and care experienced by patients has been listed as a contributory factor to disparities in cancer outcomes that exist between the UK and similar high-income countries (O'dowd, 2015).

1.1.2 Cancer pathways

Patient cancer pathways are complex systems of events and activities beginning at the point of suspected cancer and ending with patient discharge and follow-up. They are designed not only to be specific to the cancer type diagnosed but also to the subtype determined, this ensures patients receive optimum care through their cancer pathway.

The general cancer pathway itself can be broken down into several components, being made up of the referral, diagnostic and treatment pathways as illustrated in figure 1-1. Within each of these there can exist inherent weaknesses instigated by either the patient or the healthcare system itself that render the cancer pathway susceptible to delays. Understanding the points of delay and their primary stimuli within the context of the English healthcare system is the initial step in overcoming delays within cancer pathways and bringing outcomes in England on a par with those of the rest of Europe.

Figure 1-1 General cancer pathway



1.1.2.1 Primary care

The referral component of the cancer pathway represents activities taking place in the community setting prior to patient presentation in secondary care. Patients present in secondary care for investigation of suspected cancer via more than one route, these can be divided into three categories:

1. Symptomatic referrals through non-emergency routes such as those initiated by GPs
2. Referrals made via the emergency route
3. Non-symptomatic cases referred through screening programmes such as the National Health Service (NHS) breast or bowel screening programme (NCIN, 2013)

Elliss-Brookes et al. (2012) analysed data on the route to diagnosis for cancer patients in England between 2006 and 2008, they found 76% of cancers to be referred through the non-emergency route and 24% to present for further investigation in secondary care as an emergency case.

Emergency presentations of suspected cancer cases are associated with poorer survival outcomes, 1-year survival is approximately 25% lower for these patients when compared to non-emergency presentations (McPhail et al., 2013). For example, oesophago-gastric cancer patients in England were given a curative treatment plan in 16% of cases presenting through the emergency route compared to 36% of non-emergency cases and 1year survival stood at 27% compared to 43% in non-emergency cases (Palser et al., 2013). Worsened outcomes are in large part due to the relationship between the referral route and stage at diagnosis, emergency presentations are associated with a more advanced cancer stage at diagnosis (Carneiro et al., 2016) indicating a delay in the patient arriving in secondary care for investigation. Patients that present with early stage cancer at grade 1 or 2 are more likely to survive since the cancer is less likely to have spread (Hiom, 2015), the chances of survival shrinks with each increase in stage especially for lung, ovarian and bladder cancer (Public Health England, 2016). Furthermore, variances in stage at diagnosis partly explains international differences in survival for many cancers, including for lung cancer (Walters et al., 2013) colorectal cancer (Maringe et al., 2013) and breast cancer (Muller et al., 2018).

Several factors contribute to the occurrence of emergency presentations; patients delaying visiting their GP either due to lack of awareness of the signs and symptoms of cancer or barriers to access, where patients do make GP contact, many still arrive in secondary care through the emergency route while awaiting referral and others represent missed opportunities for referral by GPs (Murchie et al., 2017). Interventions designed to overcome

these and reduce the number of emergency presentations, thus the earlier diagnosis of cancers will inevitably have a positive impact on patient survival outcomes bringing them in-line with their European counterparts. Laudicella et al. (2018) explored the impacts of rerouting patients from emergency to non-emergency referral routes, they estimate over a 5-year follow-up period from diagnosis, 1863 life years will be saved for colorectal cancer cases, 889 life years for breast cancers, 1195 life years for prostate cancers and 1011 life years saved for lung cancer patients. GPs are recommended to use cancer risk prediction tools to assist referral decisions made, to increase their awareness of recurrent symptoms in the cases of delayed cancer diagnosis (Morgan and Wilkes, 2017) and to improve their adherence to referral guidelines (Car et al., 2016), recognising and referring suspected cancer cases in a timely manner for investigation and treatment in secondary care will help overcome some of the delays that generate poor cancer outcomes.

1.1.2.2 Secondary care

Once cancer patients are referred, the remainder of activities making up both the diagnostic and treatment pathways take place in secondary care. Delays occur along the diagnostic pathway, mainly instigated by weaknesses in health service infrastructure rather than by patients themselves, since at this point patients are aware of their suspected cancer status and in general comply with the urgency of the tests and investigations being carried out. Delays in diagnosis in secondary care could be driven by:

- Staff shortages within the medical departments involved in the cancer pathway
- Unavailable equipment
- Repeat investigation or examination owed to uncertain results or equipment failure
- The need to seek external expert opinions for problematic cases where the appropriate expertise is not available onsite

Delays in the diagnostic pathway from presentation in secondary care to diagnosis are however shorter than those manifesting in the primary care stage of the cancer pathway (Allgar and Neal, 2005). Delays in primary care occur at a stage where cancer may not yet be suspected or even considered, however in secondary care, cancer is evidently suspected and with it comes the urgency of confirmation and treatment initiation. Nonetheless, secondary care delays contribute to overall diagnostic delay and their removal allows cancers to be diagnosed sooner and for patients to be set on their treatment pathway earlier.

At the crossroads of the diagnostic and treatment pathway, decisions are made by clinicians on the treatment plans to follow for optimum patient care. These decisions are largely based

on the underlying characteristics of the disease reported, determined through investigation in the pathology lab within the diagnostic pathway. The disease characteristics reported that inform decision-making must be reflective of the true disease state thus allowing clinicians to design treatment plans specific to the patient needs. Published guidance based on research evidence act as the stepping-stone between diagnosis and treatment, they help guide clinicians on the treatment to recommend based on the disease characteristics reported. International variations exist in guidelines for cancer treatment but it is not known to what extent these contribute to the variations in cancer outcomes (Norell et al., 2018). At best, hospitals in England must have systems in place to ensure compliance with UK standards. Patient, primary care, referral and secondary care delays can all contribute to overall diagnostic delay for cancer patients (Allgar and Neal, 2005). However the significance of each of these by cancer type is still not yet known (Neal et al., 2015), but it is established with certainty that delays should be avoided where possible (Hansen et al., 2011).

1.1.3 Government policy

Health service systems such as the NHS have an important role to play in eliminating delays in diagnosis and enhancing treatment management for cancer patients. At both the primary and secondary care level, clinicians must work to reduce unnecessary time delays by adhering to published cancer referral and cancer management guidelines. Interventions to overcome delays are certainly welcome but their impact must be measured against agreed benchmarks to determine if they are indeed of value in overcoming the delays existing in the cancer pathway. Universal standards are set and agreed upon at the governmental level, against these the time to diagnosis and treatment across different NHS providers in England can be measured and compared. Policy set by government ensures a minimum level of care is maintained across all English healthcare providers.

Government strategies have continually kept early diagnosis at the core of their objectives for cancer care by setting minimum waiting time targets for cancer patients. These have evolved over the years as follows:

1. *NHS Cancer Plan 2000* (Department of Health, 2000)

The following targets apply to symptomatic patients presenting via urgent GP referral

- a. Maximum 2 week wait for urgent outpatient appointment in secondary care
 - b. Maximum 1 month (31 days) from diagnosis to treatment
 - c. Maximum 2 months (62 days) from urgent referral to treatment
2. *Cancer Reform Strategy* (Department of Health, 2007)

Identical targets to those outlined in the NHS Cancer plan remained in effect but they were updated as follows:

- a. The 62 day target was extended to also apply to patients arriving via the screening and the emergency route
- b. The 31 day target was extended to cover the second and third treatments in the cancer pathway where these are recommended

3. *Improving outcomes a strategy for cancer* (Department of Health, 2011)

No change from previous years

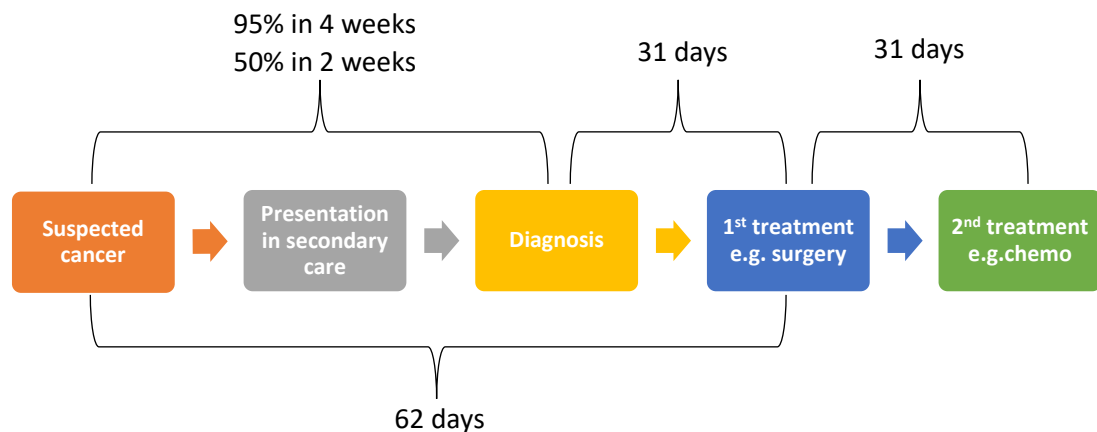
4. *Achieving world class cancer outcomes* (Independent Cancer Taskforce, 2015)

The waiting time targets of 31 and 62 days described in previous government strategies continue to stand, however additional recommendations were made for the waiting time to diagnosis, where cancer is either confirmed or excluded and the patient is discharged back into primary care. So far, government strategies have focused on time to treatment and time to referral but no specific target existed for time to diagnosis covering the investigative steps in secondary care, this phase is so far only covered by the 62- day target. The Independent Cancer Taskforce (2015) go a step further than previous recommendations, they propose a four-week target in the investigative stages to be met for 95% of suspected cancer cases by 2020, this is the period from GP referral to confirmed diagnosis or exclusion of cancer and within this time the result must be communicated to the patient. They go even further by suggesting that for 50% of cases, a diagnosis should be achieved within two weeks. With these updated recommendations, the two-week target from referral to presentations in secondary care is no longer valid, since it is now encompassed within the new four-week target. Figure 1-2 summarises the current status of cancer waiting time targets as set by NHS England.

The importance of early cancer diagnosis is unquestionable; cancers are more likely to be identified in their early stages of development with the possibility of treatment being initiated sooner. Patients have a greater chance of being recommended a curative treatment plan versus palliative care and thus survival outcomes for cancer patients are improved. Evidence of time savings in the cancer pathway have emerged since waiting time targets have come into effect. Neal et al. (2014) report a reduction of approximately five days in the time from the first symptom to diagnosis.

However early diagnosis alone is not sufficient, as patients are diagnosed they must receive the highest quality standard of treatment that is specific to their needs to maintain acceptable cancer survival outcomes. Early diagnosis is of little value if appropriate treatment management does not follow it. Cancer treatment plans are in most part governed by the diagnosis made during the investigative stages of the cancer pathway.

Figure 1-2 Cancer waiting time targets



Cancers that develop in the same part of the body can differ; appreciating the details of the cancer subtypes allows a targeted treatment approach to be taken by clinicians. Drugs are prescribed where there is a known likelihood of response and avoided together with their side effects where the feature of the cancer targeted by the drug is not present. Patients are given treatment that is expected to have an impact on their survival outcomes and avoid unnecessary treatment.

'The right treatment for the right patient at the right time' (Jackson and Chester, 2015) defines personalised medicine. This is governed by the ongoing discovery of biomarkers with the simultaneous rapid development of diagnostic tests by pharmaceutical companies (Drucker and Krapfenbauer, 2013). Test results are used to guide the decision-making process. In addition to early cancer diagnosis, treatment and molecular diagnostics were at the heart of the recommendations for improving survival outcomes in the latest government strategy outlined by the Independent Cancer Taskforce (2015). Improving access to molecular diagnostic tests for all cancer types by 2020 was a target.

While the existence of targets and recommendations is vital to measure against the performance of NHS care providers, systems must be in place to support clinicians in achieving these targets. Inputs into the cancer pathway through intervention or adaptation of the system will undoubtedly influence the performance output of NHS care providers.

Without these, government targets will only manifest as a burden, adding further pressure onto an already strained system.

1.2 Pathology

At the heart of all diagnosis lies pathology, a vital yet overlooked component of cancer pathways. Manipulation of pathology systems has the ability to contribute to improvements in both the time to diagnosis and treatment management. '*Pathology is the bridge between science and medicine*' (RCPATH, 2018c), it plays a role in all disease areas through the examination of an extensive range of bodily material including tissue specimens, blood, bone marrow and other body fluids. The material is analysed to understand the underlying cause and characteristics of disease. Clinicians use information from pathology to guide disease management including both diagnosis and treatment. Pathology services are the backbone of patient care, 70-80% of decisions related to diagnosis and treatment are founded on results from the pathology lab (Beastall, 2008).

Pathology itself covers a wide range of disciplines subject to the material being examined (Mahfouz, 2009). The Royal College of Pathologists list five main and nine smaller specialties of pathology. These are histopathology, medical microbiology, haematology, clinical biochemistry and immunology as the main branches of pathology in addition to the smaller specialties of genetics, neuropathology, cytology, clinical embryology, dermatopathology, forensic pathology, toxicology, veterinary pathology and virology (RCPATH, 2018c). Each of these covers a wide range of tests and procedures carried out by doctors and scientists in the lab in order to study and understand disease.

Histopathology is by far the most significant of all the pathology specialties, it involves the examination of tissue material under a microscope, making comparisons against healthy tissue to detect differences (Slaoui and Fiette, 2011). A specimen is removed from the patient in most cases through biopsy, it is gross examined, processed, embedded in paraffin, cut into thin sections and stained ready for microscopic examination by the histopathologist who deliver a diagnosis of the disease (Kapila et al., 2016). Histopathologists work closely with other clinicians with whom they share their findings (Masood and Horsfield, 2016). For most cancer cases, histopathologists are the primary pathology clinicians that tissue specimens are exposed to.

Pathology services are the gold standard for cancer diagnosis (Leong and Zhuang, 2011). Clinicians cannot plan or initiate treatment without the pathologists report on the clinical findings of the tissue removed from the suspected cancer by biopsy (Kufe et al., 2003).

Pathology services are located in the investigative stages of the cancer pathway in particular between the patient first presenting in secondary care and diagnosis of disease. Over recent years, with the advance of personalised medicine, the role of the pathologists in cancer diagnosis and treatment has changed dramatically (Misialek, 2014). They are often referred to as diagnostic oncologists due to the breadth and extent of diagnostic and prognostic molecular biomarkers (Srigley, 2009) they must test for to offer patients a care plan personalised to their specific disease characteristics.

1.2.1 Challenges in pathology

The continual discovery of cancer biomarkers and the introduction of their corresponding tests into patient cancer pathways requires the pathologist to first confirm the presence of malignancy as well as subtype the tumour and carry out molecular testing (Davidson et al., 2013) for targeted therapies. This increases pathology workload when diagnosing cancers, in theory amplifying diagnostic delays as a result of the extent of information pathologists must now report prior to cancer treatment initiation. Simultaneously NHS cancer targets are continually being reviewed and updated to reflect more stringent criteria consequently adding further strain on laboratories when handling the changes in demand for cancer diagnostics.

This comes at a time where pathology labs in general are being challenged due to the increasing pressure on their services. There has been an annual 4.5% increase in requests from pathology labs from a workforce that is growing at a much slower rate than for other specialities (Cancer Research UK, 2016b). Not enough medical trainees are starting pathology training and more than 30% of current pathology consultants are aged 55 or over (Cancer Research UK, 2016b), pathology labs are at the brink of a workforce crisis so that it becomes increasingly difficult to cope with the continual growth in demand for their services.

Since 2010, there has been a 17% annual increase in patients waiting more than six weeks for cancer diagnosis (Cancer Research UK, 2016b), falling behind on government targets. The rising demand for more complex specimen analysis, increased workload and a shortage of pathologists (Williams et al., 2017) have together created an environment predicted to hinder patient care through delays in diagnosis (Child and Gupta, 2009) if not confronted head on. Here we have a scenario of evident deficiencies in pathology labs further augmented by external pressures.

1.2.2 Managing demand

Pathology labs must adapt to handle the additional diagnostic pressure without undermining the ability to adhere to cancer waiting time targets or quality standards. The appropriate infrastructure and procedures must be in place to support laboratory staff, including pathologists, in meeting the increasing demands for their services. Without intervention, pathology reporting on cancer specimens will become a lengthier process and cancer waiting time targets become difficult to attain, having a knock-on effect on patients' treatment initiation and thus survival outcomes and reflecting negatively on NHS providers.

In 1999 the *Pathology Modernisation Programme* was launched by the Department of Health at the time, with the aim of improving the quality and efficiency of pathology services (Beastall, 2008). One of the first policy documents, *Modernising pathology services*, resulting through consultation as part of the programme highlighted five key areas for the modernisation of pathology in order to meet increasing demands. These were:

- Integrating pathology to support wider services
- Re-designing the systems
- The use of effective and appropriate testing
- Decreasing inappropriate variation in pathology practice across the country
- Taking advantage of new technology and IT (Department of Health, 2004)

The programme introduced the concept of new technologies and re-configuration of practice and systems in pathology to deliver high quality care for patients. The *Pathology Modernisation Programme* went on to commission two independent reviews into NHS pathology services in England. The first by Lord Carter (2006) recognised the importance of improving the quality of pathology services in the wider context of the patient pathways and recommended adapting systems and practice within pathology as well as the faster adoption of new technologies to achieve this. The second phase of the review also by Lord Carter (2008) once more recommended the adoption of innovation in both technology and practice to achieve clinical excellence in pathology.

1.2.3 Pathology innovation

Innovation in pathology is largely driven by the fundamental need to deliver high quality services efficiently within the lab to maximise the clinical benefits to the patient (RCPATH, 2019). There are currently two layers of innovation, the first is around the discovery of new tests to particularly focus and personalise treatment to the patient's needs. The modernisation of practice through the availability of more accurate and novel diagnostic

information, adds to the oncologist's ability to plan and target treatment appropriately, translating into better quality of care for the patient. This layer increases the workload of the pathology lab.

The second layer of innovations are intended to counteract the workload impacts of the first. They are centred on the introduction of modifications in the way things are done in the lab. For this group of innovations, rather than introducing an intervention that provides more information about the characteristics of a disease, they are centred on maximising the capacity of the procedures already in place for example by enhancing existing systems through technology or re-configuring current pathology workflow to enhance efficiency within the lab or modifying strategies to testing rather than the introduction of a new test. Improvements in cancer outcomes will transpire as a result of productivity gains in the lab, shrinking the time to diagnosis and enabling a greater number of tissue specimens to be processed and analysed in the given time thus contributing to early diagnosis.

Cancer Research UK (2016b) lay out a series of recommendations to address current issues within the pathology lab, one of which is digitisation. This is expected to upgrade workflows within the lab for a smoother transition from tissue specimen arrival to report ready by pathologist, in theory reducing the time taken to diagnosis and treatment initiation. Pathology activities comes under the new 28 day waiting time target, recommended by the Independent Cancer Taskforce (2015). Modernisation of pathology services that lead to time savings in this phase of the pathway are likely to have implications for overall cancer waiting times and thus improve survival outcomes. Furthermore, the digitisation of pathology is set to revolutionise the methods for analysis and reporting on tissue specimens through the introduction of computer software that allows more accurate quantification of the characteristics of disease.

NHS England (2014b)'s report on the *Digital First* pathology programme described the ways in which pathology labs were already incorporating digital methods into their practice. These include:

1. The use of a single laboratory system to consolidate pathology services across more than one site at Leicestershire Pathology Service
2. The development of an electronic consultation service between primary and secondary care, including the electronic transfer of patient records in Bradford
3. An online portal for patients to access test results as well as advice for the management of their condition

4. Using business intelligence tools such as those adopted by Sandwell and West Birmingham Hospitals NHS Trust to analyse data and better manage pathology services
5. An automatic identification system between laboratory specimens and patients set up at Oxford University Hospitals NHS trust for safer sample management (NHS England, 2014b)

Elsewhere the published literature has put forward other ideas for innovative pathology interventions that can contribute to improvements in patient pathways, including telepathology, virtual microscopy and whole slide imaging (Weinstein et al., 2009). Telepathology is the practice of pathology from a distance (Farahani and Pantanowitz, 2015), video imaging and telecommunications are used to aid diagnosis and decision-making (Weinstein et al., 2001). Virtual microscopy is the use of digital slides rather than glass slides to examine tissue specimens (Kumar et al., 2004). Glass slides are scanned to create digital files, resulting in larger digital images for the pathologist to view (Farahani et al., 2015). Whole slide imaging is the production of these digital images and viewing them using a virtual slide viewer (Weinstein et al., 2009).

New tests and innovations around their application only impact on the specific patients they target. Whilst pathology modernisation in systems and processes are predicted to have positive effects for many patients, these types of pathology interventions are expected to cross multiple pathways of different disease areas so that many patients to various degrees will feel their benefits.

1.3 Economic evaluations

Interventions in pathology driven by modernisation and innovation are certainly welcome especially where they lead to enhanced outcomes for cancer patients and contribute to reducing delays in diagnosis. However, they no doubt come at an additional cost to the NHS since in most cases they may be designed to optimise rather than replace existing procedures or they may signify an additional step in the pathology workflow. In an NHS with limited resources it is essential that proposals for investment to improve pathology services are evaluated to quantify their economic implications alongside their clinical impacts.

Economic evaluations support decision making in healthcare (Robinson, 1993e). The costs and outcomes of two interventions, in this case pathology innovation and pathology standard practice, are identified, measured, valued and compared against each other to place a value on the change in costs and benefits that will come with the implementation of

modernising interventions in pathology. There are five different approaches to economic evaluations. They all use similar methods to value the costs for comparison; the types of costs included in the evaluation must be relevant to the study perspective adopted. Economic evaluations differ in the methods they use to measure and value benefits. The different types of economic evaluations are listed:

1. Cost utility analysis (CUA): outcomes are measured using preference-based or utility-based measures such as the quality-adjusted life year (QALY) (Robinson, 1993c)
2. Cost-effectiveness analysis (CEA): outcomes are measured in natural units depending on the objective of the intervention, for example life years gained (Robinson, 1993b)
3. Cost-benefit analysis (CBA): outcomes are assigned a monetary value (Robinson, 1993a)
4. Cost-consequence analysis (CCA): More than one outcome of the intervention, not necessarily in the same unit are listed in a disaggregated format (Hoomans and Severens, 2014)
5. Cost-minimisation analysis (CMA): only costs are compared in this type of evaluation. This method is only used when it can be certain that the interventions being compared are equivalent (Robinson, 1993d)

The differences between two interventions in a CUA and CEA are conveyed as a ratio of change in cost per unit change in outcome (Garber and Phelps, 1997) referred to as an Incremental Cost Effectiveness Ratio (ICER). The National Institute for Health and Care Excellence (NICE) has set a threshold of £20,000 to £30,000 per quality-adjusted life-year (QALY) to deem an intervention cost-effective and support its introduction (McCabe et al., 2008). CBA results are expressed in Net Monetary Benefit (NMB), this is the monetary value of benefit less the cost of treatment, compared across the two interventions. Economic evaluations are central to investment decisions within a cash-strapped NHS system. Results from these support or challenge the case for the introduction of new technologies.

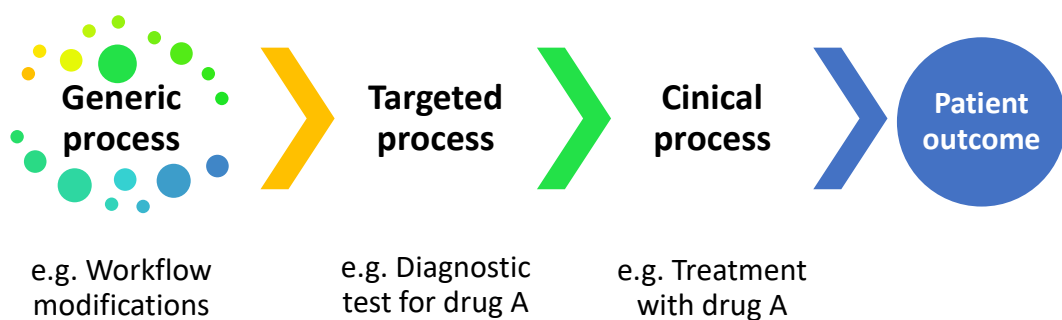
1.3.1 Challenges in the evaluation of pathology services

Carrying out economic evaluations of pathology interventions can be challenging owing to the relationship between pathology and the patient. Pathology interventions act through one or more intermediate processes before their impacts are experienced by the patient (Sutton et al., 2018). Unlike clinical interventions for example, where the effect of treatment impacts the patient directly so that clinical benefits are attributed to the intervention of interest, the output of a pathology intervention influences an intermediate process which in

turn impacts clinical outcomes. This applies to all pathology interventions whether they involve the introduction of new diagnostic tests or alterations in systems and processes. An additional layer of complexity exists when evaluating the latter type of innovation, diagnostic tests are in most cases specific to a disease area whereas innovations in systems and processes can have many effects across a wide range of patient outcomes and care pathways (Sutton et al., 2018), the benefits of system changes can be substantial when measured appropriately.

Lilford et al. (2010) summarised and presented a similar relationship between patient outcomes and policy and service interventions. Figure 1-3 adapts their description of the chain of events to pathology. To carry out an economic evaluation it is important to first understand how the pathology intervention fits into this chain of events, is it a targeted process such as a diagnostic test or a generic process such as modifications in workflow. There are no clinical processes within pathology.

Figure 1-3 Relationship between pathology processes and patient outcomes



Adapted from Lilford et al. (2010)

The overall aim of the research presented in the following thesis is to establish the cost-effectiveness of innovation in pathology, particularly focusing on generic processes and strategies introduced to modify current practice and systems or that introduce technology to support the pathology lab in handling greater demand whilst maintaining a high quality service. The thesis will focus on outcomes for cancer patients and explore how pathology innovation can play a role in early diagnosis and better treatment selection. The underlying objectives and methodology used to achieve this are detailed throughout this thesis at the start of each chapter.

Chapter 2: Economic evaluations of pathology innovations - a systematic review

2.1 Introduction

The motivation for innovation in pathology is largely driven by the increasing pressure on pathology services to meet increasing demand for more complex testing and diagnosis in an environment where workforces are shrinking and NHS services are subject to ever more stringent government waiting time targets. Pathology labs must modernise and adopt new ways for working in addition to introducing technologies into their systems to support pathologists' in order to maximise efficiency, improve accuracy and cope with increased demand whilst allowing pathologists to maintain a high quality of care service.

2.1.1 Economic evaluations within pathology

There are two ongoing developments within pathology, the first is concerned with the discovery of new biomarker and genetic tests to achieve personalised treatment, the second is concerned with the modification of systems and processes to maximise efficiency, improve productivity and enhance accuracy within the lab. The extent of economic evaluations differs between these two groups. For the development of pathology tests there appears to be an abundance of research around the cost-effectiveness of these. Two systematic reviews of pathology tests are available. Fang et al. (2011) specifically reviewed the literature for CUAs of laboratory tests published up to 2008. Their study identified 141 published studies of which 42, the largest group were CUAs of haematology/oncology¹ tests. A more recent review by Watts et al. (2017) identified 356 economic evaluations of pathology tests published between 2010 and 2015 of which 108 were tests used within haematology/oncology.

When focusing on pathology tests that specifically target cancer patients, there are also a number of published studies that evaluate the cost-effectiveness of cancer tests, including both single studies and synthesised reviews. For example, Hornberger et al. (2012) carried out a systematic review of tests to determine risk of recurrence for early breast cancer patients, within this they analysed the economic data. Eleven out of the 56 studies eligible for inclusion in their review included an economic evaluation. Huxley et al. (2015) carry out an economic evaluation of a molecular test to identify malignancy in the lymph nodes of breast cancer patients. Havrilesky et al. (2013) carry out a CEA of a test to measure

¹ Haematology/oncology is a combined speciality within the US

metastases in endometrial cancer. D'Andrea et al. (2015) carry out a review of economic evaluations of genetic tests used in the cancer pathway.

However the scope of economic evaluations of innovations in pathology that disrupt current processes, existing systems or strategies of work to introduce new ways of doing things in the pathology lab in order to cope with demand or improve the accuracy of existing procedures appear to be a lot more limited. These are important sources of information for decision makers. Process modifications are expected to be of considerable benefit due to the extended impact they can have for many disease areas, as a result, the cost per patient is likely to be minimal, but in most cases they will involve a considerable amount of upfront and on-going investment. Economic evaluations can be used to justify and encourage the adoption of such processes. These are especially important within the cancer pathways as they can contribute to achieving waiting time targets and improving patient outcomes.

PROSPERO, The Cochrane Database of systematic reviews, TRIP medical database and the ARIF reviews database were all searched, no completed or ongoing systematic reviews of economic evaluations of innovation in pathology were identified.

2.2 Aims

The aims of this chapter are as follows:

1. To systematically review and summarise the methods applied when carrying out economic evaluations of histopathology process interventions through their impact on cancer outcomes.
2. To summarise the cost-effectiveness evidence on innovations in pathology services for cancer patients

2.3 Methods

The Centre for Reviews and Disseminations guidance for carrying out reviews in healthcare (CRD, 2009) were followed to guide the review process.

2.3.1 Information sources

A broad literature search was carried out to identify all the published economic evaluations of innovation in pathology. Both bibliographic database searches and forensic searching techniques were carried out to identify all the relevant literature.

2.3.1.1 Electronic searches

The following databases were searched for full economic evaluations:

1. DARE
2. Econlit
3. Embase
4. HTA
5. Medline
6. NHS EED
7. Web of Science

The search strategies used in the database searches combined three key concepts, in essence including key word and MESH search terms that describe cancer, pathology and economic evaluations. To maintain an inclusive approach to the types of patient outcomes that can be measured and the comparators used, no search terms related to either of these concepts were included in the search strategies so that the full breadth of economic evaluations are captured. Furthermore, whilst innovation is a clear criteria for inclusion of a publication, associated terms were not built-in so as not to restrict the search results, instead a year limit of 2000 was set, this ensures that the maximum number of studies are captured and those that were not relevant were excluded at the screening stage of the review. The search strategy was first developed in Medline and adapted for use in the remaining databases listed. The search strategies for each of the databases are given in appendix A.

2.3.1.2 Other sources

Forward and backward citation was carried out on all the included studies identified through the electronic searches to locate any further research not picked up in the initial database searches. In addition to this, key word searches were carried out in Google (and Google Scholar) and the first three pages of search results screened against the eligibility criteria.

2.3.2 Eligibility criteria

2.3.2.1 Types of studies

All types of full economic evaluations that measure the difference in both costs and outcomes of two or more alternative strategies were eligible for inclusion in the review. These include cost-effectiveness analyses, cost-utility analyses, cost-consequence analyses, cost-minimisation analyses and cost-benefit analyses. They include economic evaluations that are carried out alongside randomised controlled trials (RCT) or those that use decision-analytical modelling techniques as a vehicle for the analysis. Economic analyses that do not meet the requirements of a full economic evaluation were not eligible for inclusion in the review. All types of partial economic evaluations as listed in table 2-1 were excluded during the screening stages of the review but their total number and types of interventions that were evaluated was recorded.

All conference proceedings were excluded, however they were initially screened against the eligibility criteria. Where they referred to a full economic evaluation, the corresponding full paper was searched, for inclusion in the review if it had not already been identified through the electronic searches.

2.3.2.2 Types of participants

Since pathology services play a vital role in both the diagnosis and management of cancer cases, two groups of patients were eligible for inclusion in the review. All studies where patients with suspected cancer are the population of interest and secondly all studies where the population of interest had been previously diagnosed with cancer. There were no restrictions on the types of suspected or diagnosed cancers and no further criteria concerning the types of participants. Full economic evaluations that explore the cost-effectiveness of pathology interventions targeting these groups of patients were eligible for inclusion in the review.

2.3.2.3 Types of interventions

Pathology interventions designed to disrupt current systems within the pathology lab either through the introduction of a new technology or the modification of current processes to support growing demand on pathology services or to improve the accuracy of existing tests were eligible for inclusion in the review. Due to the breadth of pathology, interventions had to particularly target the handling of tissue specimens (histopathology) to be included. Economic evaluations of newly discovered biomarkers and genetic tests that have not previously been adopted were excluded however, economic evaluations of strategies for

testing for molecular biomarkers that have already been approved and are in use were included.

2.3.2.4 Types of comparators

All studies that were eligible for inclusion in the review, had to include at least two arms in their economic evaluation. The first representing the pathology innovation being evaluated and the other(s) a suitable comparator. If an appropriate comparator could not be identified by the review authors they had to include a 'do nothing' arm for their study to be eligible for inclusion in the review. Studies that do not include a comparator arm were excluded from the review.

2.3.2.5 Types of outcome measures

Pathology services are classified as indirect interventions to patient care and their effects are not always measured in relation to patients. Economic evaluations had to explicitly keep cancer patients at the heart of the analysis and measure the effect of the intervention directly to them to be eligible for inclusion in the review. Beyond this requirement there were no restrictions on the types of outcomes that an economic evaluation could use to measure the effect of the pathology intervention.

Table 2-1 Summary inclusion criteria

	Include	Exclude
Population	All patients that are suspected or have been diagnosed with any cancer	Conditions other than cancer
Intervention	Histopathology interventions that include: <ul style="list-style-type: none"> • New technologies • Modifications in systems and processes • Alternative testing strategies 	Non-pathology interventions Pathology interventions that are not specific to histopathology Newly discovered molecular and genetic tests
Comparator	Any comparator	No comparator
Outcome	Any outcome but must be related to cancer-patient	Outcomes not related to cancer patient
Study Design	All types of full economic evaluations: <ul style="list-style-type: none"> • Cost-benefit analysis • Cost-consequence analysis 	All types of partial economic evaluations: <ul style="list-style-type: none"> • Cost analysis • Cost description

	<ul style="list-style-type: none"> • Cost-effectiveness analysis • Cost-minimisation analysis • Cost-utility analysis 	<ul style="list-style-type: none"> • Outcomes analysis • Outcomes description
Country of Research	Europe and North America	Countries outside of Europe and North America
Language	English only publications	Non-English publications
Publication Date	2000 Onwards	Pre-2000

2.3.3 Data collection and analysis

Search results from the electronic databases as well as other sources were uploaded and managed in the reference manager EndNote. Duplicates were removed.

2.3.3.1 Study selection

Two reviewers were involved in deciding on the eligibility of studies for inclusion in the review. A two-step process was taken for study selection as recommended by CRD (2009) guidance. The first step involved the two reviewers independently screening the titles and abstracts of all search results against the pre-specified inclusion/exclusion criteria listed in table 2-1. An inclusive approach to study selection was taken at this stage of the screening process, only studies that were clearly irrelevant were excluded, all those where a decision could not be made with certainty were included at this stage and proceeded onto the second round of the screening process. Any disagreements between the two reviewers were mediated through consultation with a third reviewer.

The second round of screening started with the retrieval of the full papers that passed the first stage of the two-step process. The full texts were screened one more time against the criteria listed in table 2-1 and the reasons for exclusion were recorded. Where more than one published study reported on the same economic evaluation these were regarded as one piece of research. Once again, any disagreements between the two reviewers on the inclusion or exclusion of a study were mediated through consultation with a third reviewer.

Published research that passed through both stages of the screening process were included in the study. Their references were checked and citation searches carried out to identify further eligible research. All the stages of study selection were recorded in a PRISMA flow diagram.

2.3.3.2 Data extraction and management

Following the identification and retrieval of the eligible studies for inclusion in the review, the data points required from these to meet the study objectives were extracted using a pre-

designed data extraction form. This can be found in appendix B. The data extracted was recorded in an excel spreadsheet, ready for analysis. The information documented included general study details, descriptions of both the pathology intervention and the comparator, economic evaluation approach taken and all details of the methods used, the thresholds applied by the study authors to determine cost-effectiveness and finally the cost-effectiveness decision made concerning the intervention of interest.

2.3.3.3 Reporting quality

The reporting quality of the eligible studies was assessed against the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) checklist for reporting economic evaluations of healthcare interventions (Husereau et al., 2013). Each study was scored out of 24 and the results recorded in an excel spreadsheet.

2.3.4 Data synthesis

The cost effectiveness results including incremental costs and outcomes were summarised to establish the overall outcomes of economic evaluations of pathology interventions. Where possible studies were classified according to the population cancer and intervention type being evaluated to highlight the areas where economic evaluations were more prevalent and the different approaches used across each of these.

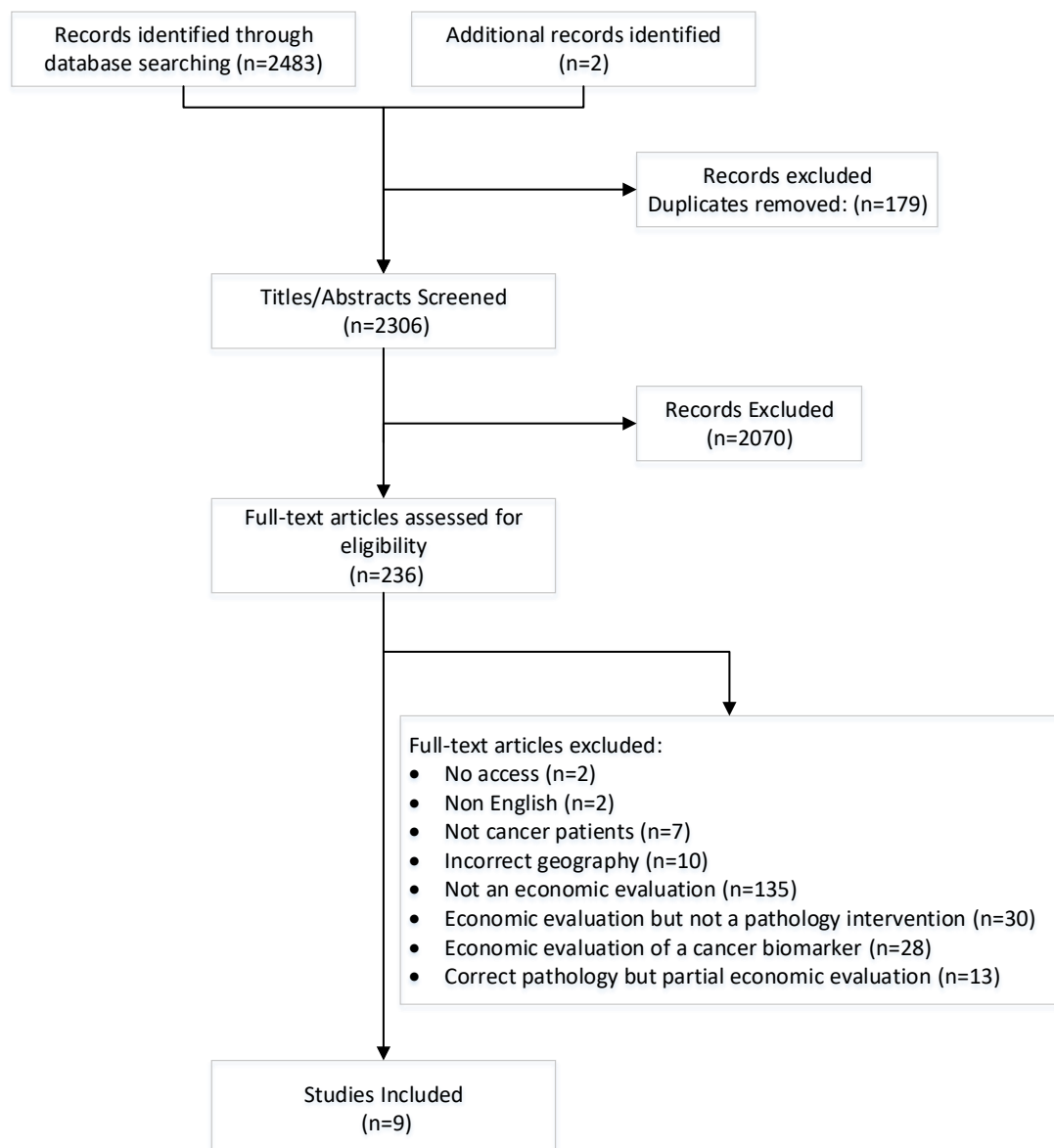
A narrative synthesis of the data collected was undertaken, through the quality assessment scores it was possible to identify and highlight the common strengths and weaknesses across the studies including lessons that can be learnt when carrying out future economic evaluations of pathology interventions.

2.4 Results

2.4.1 Search results

The systematic review database searches identified 2483 publications, after de-duplication, 2304 studies remained and were screened against the eligibility criteria. 2070 were discarded on the first round of screening, after retrieval and screening of the remaining full papers, 7 studies met the systematic review criteria. Two further studies were identified through citation and reference searching. In total 9 studies met all the eligibility criteria for inclusion in the review. A summary of the study selection process is given in figure 2-1.

Figure 2-1 Study selection process



227 studies were excluded on the second round of screening, on review of the full papers of each record, the reasons for exclusion are listed in figure 2-1. Of all the studies excluded, 28

studies were excluded because they were economic evaluations of new cancer biomarkers and 13 studies were excluded because they met the pathology intervention criteria but were not full economic evaluations. The types of interventions partially evaluated were recorded, they included:

1. Intraoperative pathology procedures
2. Alternative strategies for testing for cancer biomarkers
3. Access to second opinion pathology
4. Whole-mount processing and evaluation of specimens
5. A feedback programme for pathologists

The majority of these are targeted pathology processes that are specific to a cancer type. However, access to second opinion pathology and a feedback programme for pathologists are both generic pathology processes that will have a diffuse effect across many cancers and disease areas. Coblenz et al. (2001) compared the cost of seeking a second opinion to the cost of avoided surgery for patients with bladder carcinoma, they reported a cost saving as a result of their intervention. No patient outcomes were considered as their evaluation is a cost-cost offset analysis. Torres et al. (2003) evaluated the impact of a feedback programme for pathologists involved in reporting on breast disease. The authors presented a cost description of the intervention and did not attempt to measure these against any benefits.

2.4.2 General study characteristics

9 studies were eligible for inclusion in the review. Figures 2-2, 2-3 and 2-4 describe the general characteristics of these. 8 out of the 9 published studies investigated pathology interventions in the breast cancer pathway. The remaining study evaluated a pathology intervention introduced into the thyroid cancer pathway. The majority of studies were from the US and Canada and only 2 were European based studies. Noticeably none of the studies were UK based.

Of the pathology interventions evaluated, the majority were evaluations of alternative testing strategies for the detection of a particular biomarker. These were not evaluations of newly discovered biomarkers in the same way as for the 27 studies that were excluded from the review as described in figure 2-1. These were evaluations of biomarkers that were already proven cost-effective and were in use in the pathway. The motivation behind these evaluations was to determine the most appropriate test strategy to adopt to ensure patients receive the correct treatment. All studies in this group compared a range of testing strategies that can be used to determine the biomarker HER2 status in breast cancer. They included

Elkin et al. (2004), Garrison et al. (2013a), Blank et al. (2010), Morelle et al. (2006), Dendukuri et al. (2007) and Lee et al. (2011).

Figure 2-2 Summary of cancer types investigated

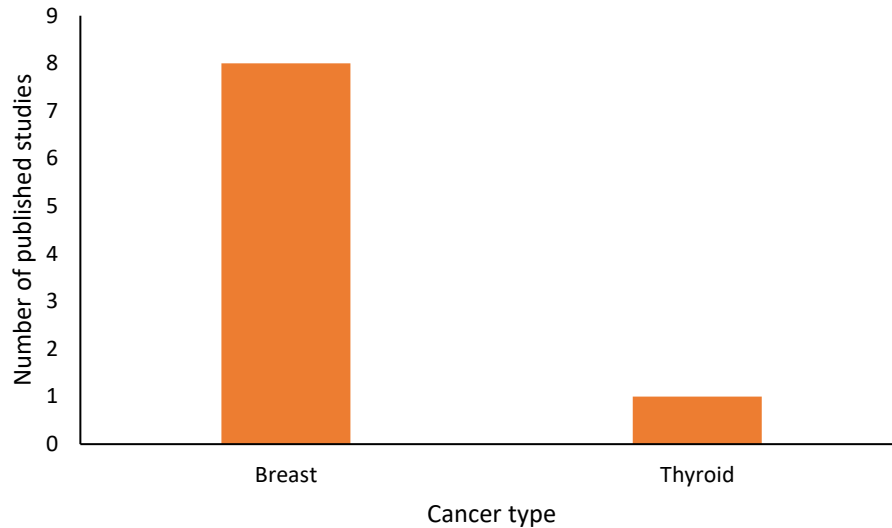
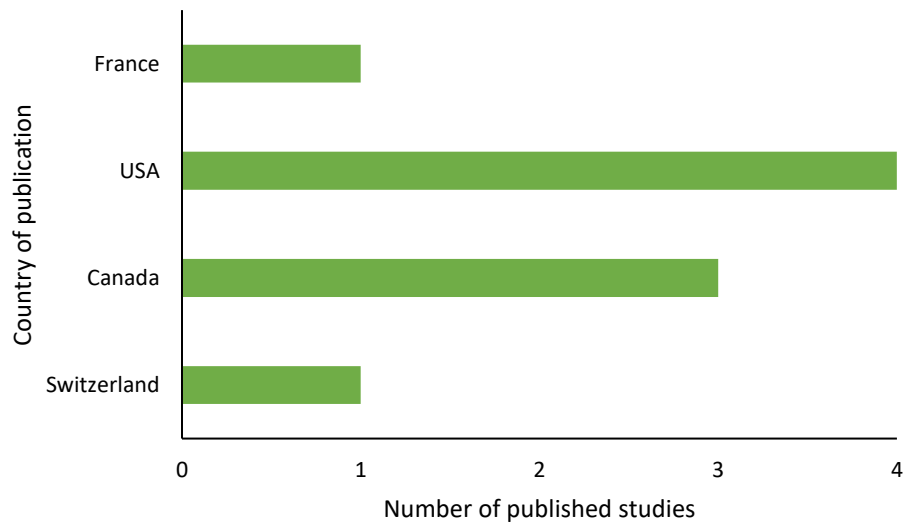


Figure 2-3 Summary of papers by country of origin

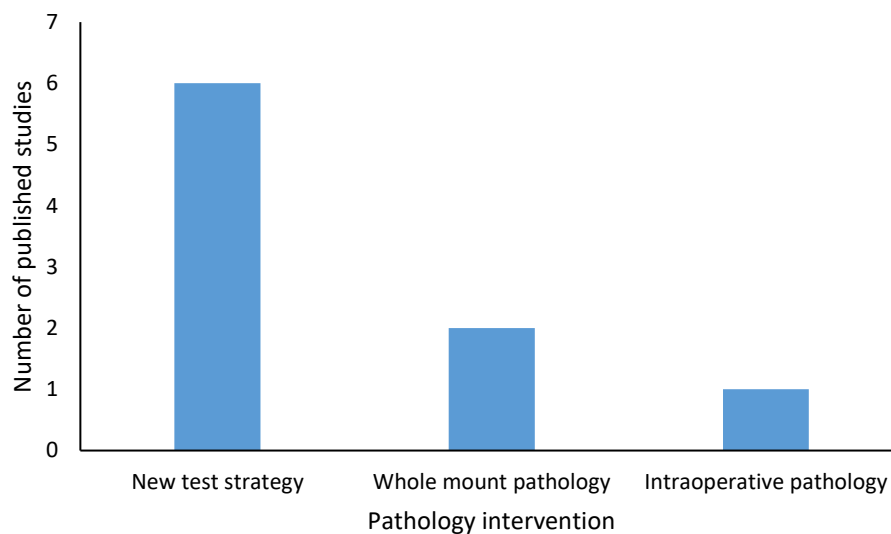


Beyond the alternative testing strategies, Zanocco et al. (2013) evaluated the cost-effectiveness of the use of intraoperative pathology examination (IPE) when investigating neoplasms in palpable thyroid nodules. Conventionally the tissue specimen is sent for pathological examination in the lab and the patient will return for second total thyroidectomy surgery if there is evidence of malignancy on pathological examination. With IPE the tissue specimen can be examined immediately within the first surgical time period so

that if second surgery is recommended it can be completed straightaway rather than delayed.

Two papers by Look Hong et al. (2016b) and Look Hong et al. (2016a) report on the final pathology intervention identified in the list of eligible studies. They both evaluate the cost-effectiveness of using whole-mount (WM) histopathology processing rather than conventional processing for the examination of breast specimens. Under the conventional pathology techniques tissue specimens are sampled onto small slides, because the whole specimen is not represented in the slides there is a risk of missing a diagnosis due to under sampling of the surgical tissue in the slides (Clarke et al., 2007). With WM histopathology 30 times more tissue than with conventional methods is evaluated due to the preservation of the 3D conformity of the specimen (Look Hong et al., 2016b). Even though they both evaluate the cost-effectiveness of the same intervention, they were both included as they adopt a different approach to the economic evaluation in their published studies.

Figure 2-4 Summary pathology interventions investigated

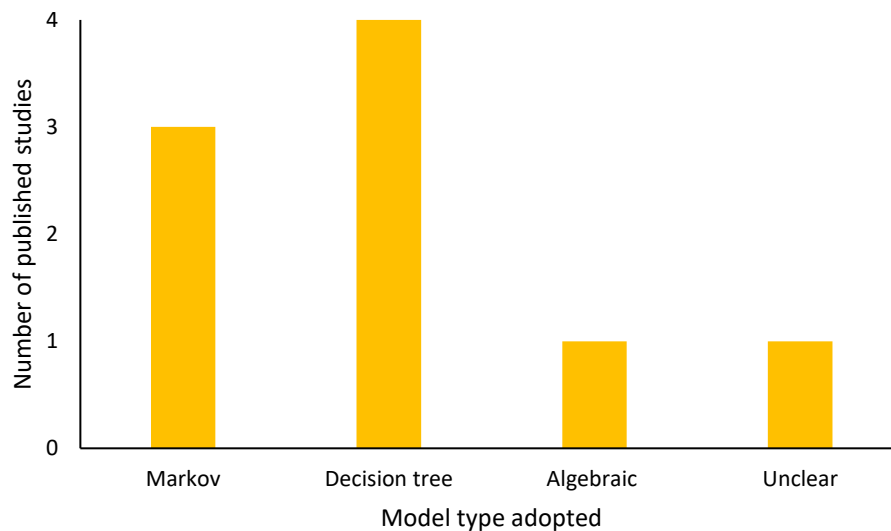


2.4.3 Economic evaluation methods adopted

The next section focuses on the methods adopted in the eligible studies to determine the cost-effectiveness of the pathology intervention of interest.

All studies eligible for inclusion in the systematic review used a form of decision modelling as the base for the economic evaluation. The majority of authors constructed either a decision tree or Markov model. A summary of the types of models adopted is given in figure 2-5.

Figure 2-5 Summary modelling approaches adopted



6 studies evaluated the cost-effectiveness of alternative strategies to determine HER2 status to guide Herceptin treatment for breast cancer patients. There are two commonly used methods to determine HER2 status, Immunohistochemistry (IHC) or fluorescence in situ hybridisation (FISH). Many of the strategies explored in these studies include either one or a combination of these tests. IHC will generate one of the following three results, 0/+1 (negative), +2 (equivocal) or +3 (positive). FISH results are either positive or negative.

Approaches to testing is a targeted process intervention in the pathology lab, it is specific to a patient group and does not have a diffuse effect. Out of the 6 studies that evaluated the HER2 test strategy, 4 mapped the targeted process onto Herceptin treatment, the clinical process, to determine QALY gains as a result of treatment. These were Blank et al. (2010), Elkin et al. (2004), Garrison et al. (2013a) and Lee et al. (2011). The remaining two studies did not measure QALYs by mapping the outcomes of testing onto Herceptin treatment. Morelle et al. (2006) worked out the cost per correctly managed breast cancer case and Dendukuri et al. (2007) measured the cost per correct determination of HER2 status.

All of the HER2 testing strategy studies used decision-modelling techniques as a vehicle for their economic evaluation. Blank et al. (2010) and Elkin et al. (2004) used Markov models to estimate costs and QALYs over the breast cancer patient's lifetime. Both studies reported all transition probabilities and cost and effectiveness parameters including their sources that were used to populate the model and provide illustrations of their model structures. Blank et al. (2010) compared 6 strategies:

1. Patients receive IHC test only
2. Patients receive FISH testing only
3. Parallel testing using both IHC and FISH tests
4. FISH testing if IHC equivocal
5. No testing and all patients receive Herceptin treatment.

The sensitivity and specificity of each test strategy were based on the published literature. A strength of this evaluation is its use of 4 independent sets of transition probabilities for each type of breast cancer patient i.e. those that are true positive, false positive, true negative and false negative to ensure that the differences in accuracy across each strategy are modelled appropriately. Blank et al. (2010) follow their base case analysis with both deterministic and probabilistic sensitivity analyses.

Elkin et al. (2004) also use a Markov decision model as a vehicle for their economic evaluation. They compare 7 strategies for HER2 testing:

1. No test and chemotherapy alone
2. IHC test and Herceptin treatment if IHC +3
3. IHC test and confirm +2 and +3 results with FISH
4. IHC test and confirm +2 results with FISH
5. IHC test and Herceptin treatment if +2 or +3
6. FISH test only and Herceptin if positive
7. No testing, Herceptin and chemotherapy treatment given

To determine the sensitivity and specificity of the IHC test they identify 10 studies that compare IHC to FISH results and use these to work out the weighted average with confidence intervals for the probability of each test score. The main concern with this approach is the absence of consideration for the underlying scoring thresholds that are used to determine positive and negative status, these have changed over time due to updated guidelines. This issue is discussed extensively in chapter 6 later on in the thesis. Furthermore, it is not clear from the study if the model transition probabilities reflect HER2 test results conditional on the true HER2 status. Unlike the transition probabilities reported by Blank et al. (2010), Elkin et al. (2004) report transition probabilities for HER2 and negative cases without subdivision by those that are false and those that are true. The results from this study may not be truly representative of the differences in accuracy with each test strategy. Elkin et al. (2004) carry out one-way and two-way sensitivity analyses to measure uncertainty around their base case analysis estimates.

Garrison et al. (2013a) and Morelle et al. (2006) also carry out economic evaluations of alternative HER2 testing strategies, they use decision tree models as the vehicle for their economic analysis. Garrison et al. (2013a) compare two testing strategies:

1. Standard practice: IHC test as first-line with repeat test on +2 scores using FISH.
2. Proposed expanded reflex testing strategy: IHC as first-line and retesting both +1 and +2 scores using the FISH test.

An illustration of their model structure is presented in their study. The model probabilities they use are based on a single study that estimates the probability of each score for the IHC and FISH test, however they do not reflect on accuracy in their model by separating out the false and true cases through the consideration of test sensitivity and specificity. This would reflect onto the cost and QALY values used, since undoubtedly the QALY gain of patient that is a true positive is expected to be greater than the QALY gain of a breast cancer patient that is false positive. Garrison et al. (2013a) carried out both deterministic and probabilistic sensitivity analysis to assess the robustness of their results.

Morelle et al. (2006) also develop a decision tree model for their cost-effectiveness analysis however; they do not include an illustration of their model structure. They compare five strategies for the determination of HER2 status at metastatic relapse for breast cancer patients:

1. IHC at first metastatic relapse (current practice)
2. Systematic IHC at diagnosis
3. FISH at first metastatic relapse
4. IHC followed by FISH for +2 cases
5. IHC followed by FISH for +2 and +3 cases

The study authors estimated IHC sensitivity and specificity using data from a study that they have access to. They do not provide a description of the methods or primary data they use to estimate sensitivity and specificity. They also do not provide estimates of the resultant sensitivity and specificity that they calculate and input into the model. They also do not provide a clear summary of the cost and effectiveness parameters used to populate the model. They do not link test results with treatment and present their results as the cost per number of correctly managed patients. Deterministic sensitivity analysis was also carried out. It is challenging to follow this study through and understand the methods they have adopted.

Lee et al. (2011) compared the costs and benefits of five strategies to determine HER2 status. These were:

1. Primary FISH
2. Primary SISH
3. Primary CISH
4. Primary IHC with FISH to confirm equivocal cases
5. Primary IHC plus FISH to confirm equivocal and positive cases

They did not compare the five against each other but carried out three different cross comparisons:

- Compare strategies 1, 4 and 5 against each other
- FISH versus CISH
- FISH versus SISH

They develop an algebraic model in excel but no description or illustration of this is presented in the published study. The study authors identified published literature that report HER2 scores for each of the alternative tests and estimate sensitivity and specificity for each of these. Interestingly they did not attempt to pool or summarise their estimates but run the model to reflect each study they have identified. The model is run more than once for each analysis, each time reflecting accuracy estimates from a separate study. The model is run 46 times corresponding to 46 different studies that estimate sensitivity and specificity of IHC, to compare strategies 1, 4 and 5. For each version, estimates of cost-effectiveness are presented for each of the three strategies. To compare FISH and CISH the model is run 21 times each based on estimates of sensitivity and specificity from 21 different studies. For each version, 21 cost-effectiveness estimates are presented. Finally, the model is run five times to compare SISH and FISH so that five estimates of cost-effectiveness are drawn.

The final study, investigating the cost-effectiveness of alternative HER2 strategies is by Dendukuri et al. (2007). They compared 7 approaches for HER2 testing:

Table 2-2 HER2 testing strategies compared by Dendukuri et al. (2007)

Strategy	Condition for Herceptin treatment
Standard practice: IHC plus FISH if +2	+3 IHC OR positive FISH
IHC only	+2 OR +3 result
IHC only	+3 result

IHC plus FISH if +1 or +2	+3 IHC OR positive FISH
IHC plus FISH if +2 or +3	Positive FISH test
IHC plus FISH for all scores	Positive FISH test
FISH only	Positive FISH test

The study authors do not explicitly state that they used a model for their analysis however they refer to it in the description of their methods, the type of model used or a description of it is not given. They only consider the costs of diagnosis and report the ICER as the cost per accurately determined HER2 status.

Look Hong et al. (2016b) and Look Hong et al. (2016a) are both economic evaluations of the use of WM histopathology in the processing of breast cancer surgical specimens. Even though they report on the same pathology intervention, both studies were included in the systematic review as they take a different approach to the economic evaluation. Look Hong et al. (2016b) constructed a decision tree model to measure the impact of WM histopathology processing to conventional processing. They presented four outcomes in their study:

1. Proportion of patients with additional operation
2. Total number of hours needed for tissue processing
3. Total number of hours needed for pathologist reporting
4. Total number of slides processed

They modelled the sensitivity and specificity for each intervention onto repeat operation if malignancy is detected. Sensitivity and specificity estimates for both arms of the evaluation were taken from a single published study. Number of hours for tissue processing and pathologist reporting were based on observational data. Four ICERs were reported, one for each of the outcomes listed. Both deterministic and probabilistic sensitivity were carried out.

The second study by Look Hong et al. (2016a) constructed a Markov model to carry out a cost-effectiveness analysis comparing WM histopathology processing to conventional processing. Unlike the first, they estimated costs and QALYs over a 10-year time horizon. Model parameter estimates for sensitivity and specificity were identical across both studies. Both deterministic and probabilistic sensitivity analyses were carried out.

In reality, the studies build on each other's work, the first maps out the benefits of WM histopathology for intermediate endpoints such as the rate of re-operation and the second

study goes a step further by linking WM histopathology to the patient outcomes by measuring QALYs.

Zanocco et al. (2013) is the final study that was eligible for inclusion in the review. They carry out a CEA of the use of intra-operative pathology examination (IPE) to report on diagnostic hemithyroidectomies over a lifetime horizon. The authors do not explicitly define the perspective of their evaluation, however, this can be inferred from the costs included. They take a societal perspective and measure effectiveness using QALYs. Zanocco et al. (2013) construct a Markov model for the economic evaluation. They compare three surgical procedures with or without intra-operative pathology as listed:

1. Diagnostic hemithyroidectomy without IPE
2. Diagnostic hemithyroidectomy with IPE
3. Total thyroidectomy with IPE

Clear diagrams were given for each version of the model that represented the three interventions of interest.

2.4.4 Cost-effectiveness results reported

Four studies compared more than two HER2 testing strategies, Blank et al. (2010), Elkin et al. (2004), Lee et al. (2011) and Dendukuri et al. (2007). All the studies listed and compared incremental costs and benefits for their included strategies in a tabular format. Blank et al. (2010) concluded that FISH testing alone was the most cost-effective strategy with an ICER of €12,245/QALY, two of the other strategies were dominated and the remaining two had ICERS of €400,154/QALY and €13,456,577/QALY. They conclude that both these ICERs are high but do not explicitly state what the cut-off threshold is to consider an intervention cost-effective.

In the Elkin et al. (2004) study four of the strategies were dominated compared to standard practice and so not considered cost-effective. Two strategies, FISH test only and IHC with FISH test for +2 and +3 cases were the only two strategies not dominated. However, they had ICERs of \$125,100/QALY and \$145,000/QALY respectively. The study authors do acknowledge that these ICERs are high compared to other breast cancer interventions but they do not explicitly state the cut-off threshold for cost-effectiveness.

It is difficult to draw a cost effectiveness conclusion with the results presented by Lee et al. (2011) they present several tables with cost-effectiveness estimates that are based on

different values of sensitivity and specificity and do not explicitly conclude which strategy is cost-effective.

Dendukuri et al. (2007) also present their incremental costs and QALYs for each strategy in a tabulated format. Three testing strategies are not dominated and have ICERs that range from \$6175 to \$8401 per accurately determined HER2 status. Once more they do not explicitly state their cut-off threshold for cost-effectiveness.

Morelle et al. (2006) evaluated four strategies for scoring HER2 against standard practice. They presented ICERs in the form of cost per correctly managed case. However, they did not present an overall ICER but two ICERs per testing strategy. One ICER per correctly managed case where false negative results are avoided and a second ICER per correctly managed case where false positive results are avoided. For the false negative cases avoided, only one strategy was not dominated and for the false positive cases avoided, two strategies were not dominated. There was not a single strategy that was not dominated across both criteria. They conclude that standard practice is the efficient strategy when both criteria are considered.

Garrison et al. (2013b) estimate their expanded reflex testing approach would generate an additional \$39,745/QALY gain and \$36,721 per life year gained and state that is clearly below the cost-effectiveness thresholds. They also carry out a probabilistic sensitivity analysis and present a probability cost-effectiveness curve. There is an 87% chance of cost-effectiveness at the \$50,000 threshold and at the \$100,000 threshold, cost-effectiveness of their proposed strategy to testing is certain.

Zanocco et al. (2013) economic evaluation of IPE carried out alongside diagnostic hemithyroidectomies estimated that the intervention proposed was dominated by standard practice of conventional pathology. IPE was both cost increasing and QALY reducing.

Look Hong et al. (2016b) and Look Hong et al. (2016a) both evaluated the cost-effectiveness of WM histopathology for the examination of breast specimens. The first study estimated four ICERs:

1. \$9495/additional surgery
2. \$122/each additional hour spent in processing
3. \$5306/each additional hour of pathologist time saved
4. \$133/slide avoided

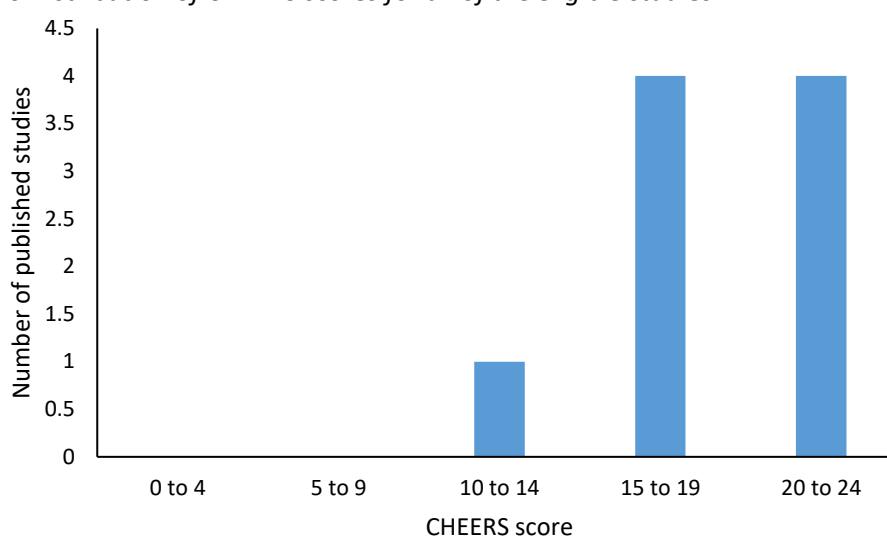
They used the results per additional surgery to construct a cost-effectiveness acceptability curve and demonstrate that their intervention is cost-effective at thresholds up to \$15,000.

The second study, (Look Hong et al., 2016a) estimated the effectiveness of WM histopathology using QALYs. They calculated an ICER of \$45414 per QALY gained. Deterministic and probabilistic sensitivity analysis supported their conclusion.

2.4.5 Quality assessment

The CHEERS checklist was used to assess the quality of all included studies. A summary of the distribution of scores for the included papers is given in figure 2-6. The average number of items that the eligible papers scored positively on was 18 and the scores ranged from 11 to 22 items completed.

Figure 2-6 Distribution of CHEERS scores for all of the eligible studies



All except one of the eligible studies, Dendukuri et al. (2007), scored positively on over 50% of the CHEERS checklist items. All the eligible studies clearly stated that a CEA was carried out and included well-structured abstracts and summarised the broader context of the evaluation in their introductory section. The study perspective was clearly defined in all except two studies, 6 adopted a healthcare system perspective and one study adopted a societal perspective. The comparators were clearly listed in all the eligible studies. The time horizon adopted was clearly defined in all except two studies, Dendukuri et al. (2007) and Elkin et al. (2004). Three studies adopted a life-time horizon, Blank et al. (2010), Garrison et al. (2013b) and Zanocco et al. (2013), three studies Lee et al. (2011), Look Hong et al. (2016a) and Morelle et al. (2006) each adopted a 15year or 20 year time horizon and finally Look Hong et al. (2016b)'s time horizon ends when the patient completes surgery. All of the studies presented their cost-effectiveness results as ICERs and carried out sensitivity analysis to confirm their findings. There is not a criterion in the CHEERS checklist that can be considered a weakness across all the studies.

2.5 Discussion and conclusion

The systematic review emphasised the lack of economic evaluations of generic pathology processes. No economic evaluation was identified that analyses the impacts of changes in systems and processes or the introduction of new technology within pathology to manage the increasing challenges that are being faced. The testing strategies investigated are targeted pathology processes specific to breast cancer. The IPE and WM are also targeted processes but the same techniques can be adopted across other cancers or disease areas where a larger specimen than what is currently possible with conventional processing needs to be examined.

The majority of the eligible studies carried out economic evaluations investigating interventions that are designed to improve the accuracy of pathology reporting. Alternative testing strategies are suggested to maximise the likelihood of patients being given the correct treatment as a result of their diagnosis and WM histopathology ensure that no malignancy on specimens is missed so that patients are considered for re-operation as appropriate. IPE is driven by its ability to reduce the number of second surgeries and patients having to return for re-operation. These types of interventions result in time savings in the management pathway to both the patient and clinicians, however these were not measured. Zanocco et al. (2013) used estimates of sensitivity and specificity and conclude that their intervention is not cost-effective.

The use of decision modelling to model the costs and benefits of the pathology interventions onto treatment selection and thus onto patient outcomes was the most commonly used approach. None of the eligible studies were RCTs where patients are randomly assigned to the intervention and comparator arms.

Pathology labs are involved in the diagnosis of disease so that the accuracy of the different pathology interventions that are being evaluated to guide treatment is at the heart of all the differences in costs and benefits measured between the intervention and its comparator. It is vital these are measured accurately and truly represent the underlying differences between the approaches under evaluation. All the eligible studies that reported accuracy, did so in the form of sensitivity and specificity estimates. Whilst most used a single pair of accuracy estimates taken from the published literature to inform their models, two studies used multiple estimates of sensitivity and specificity. Elkin et al. (2004) estimated a weighted average across more than one study for each accuracy parameter and Lee et al. (2011) rather than averaging or pooling accuracy estimates used each independent pair to inform their

model and so generated a range of cost-effectiveness results. No study used the recommended approach of meta-analysis to pool estimates from more than one study (Haidich, 2010) .

A limitation of this systematic review is its inclusion of cancer patients only as the study population of interest. A broader approach to the systematic review beyond cancer could perhaps have identified a greater number of examples of pathology innovation that have been evaluated, since pathology interventions can have far-reaching consequences. However, this would have been a very broad study and would not meet the objectives of the thesis, which is focused on cancer patients and improvements in their management through pathology interventions.

Even though the modernisation of pathology labs has been at the centre of policy recommendations since 1999, there were no UK based studies. Several pathology modernisation initiatives are currently being rolled out across England. Consolidation of pathology services through centralisation of the workforce and the use of large-capacity automated systems is a project that has been proposed and endorsed by the two Lord Carter reports (Karakusevic et al., 2016). Satta and Edmonstone (2018) have identified cost savings in laboratories that have adopted consolidation methods when compared to labs that have not. They did not explore beyond the cost implications, to measure and value changes in accuracy and time when pathologists adopt a consolidated approach to working in the lab.

A second initiative supported and endorsed by the UK government is the digitisation of pathology services. The UK government has recently invested in five new centres of excellence for digital pathology (DP) (Gov.UK, 2018) yet there are no economic evaluations of this intervention. DP is expected to have impacts on pathology workflow and accuracy especially when computer image analysis (IA) methods are used alongside. A recent survey of DP usage within the UK found approximately 60% of participating institutions to have access to a DP scanner but in most cases only applied its usage to teaching and research (Williams et al., 2018b). While many of the pathology departments agreed that the use of DP was priority within their institution, 82.5% of the respondents highlighted financial cost as a barrier to the use of DP. The institutions that responded identified data proving the intervention as cost-saving to be a strong enabler of its introduction (Williams et al., 2018b).

DP is a generic pathology process that will have wide reaching impacts across many disease areas but in particular for cancer patients, enhancements in pathology workflow will

contribute to earlier diagnosis in secondary care and improvements in accuracy owing to IA can contribute to better management of disease.

To start to put-together an economic case to justify the use of DP its impacts must be mapped onto patient outcomes through the use of economic evaluations that measure changes in both costs and benefits. Due to the scope of the impacts of DP, this can only be done within one disease area at a time. The focus of the thesis is on cancer patients, but even within cancer there are many management and treatment pathways that maybe influenced.

The majority of the published research identified and included in the study was related to breast cancer treatment and diagnosis. Breast cancer accounts for 15% of cancers in the UK (Cancer Research UK, 2015). It is the most common cancer and one of the most funded cancer research areas (Carter and Nguyen, 2012), it bears one of the largest burdens on society. Any interventions such as DP that can contribute to improving outcomes for this group of patients is welcome.

In conclusion, the systematic review highlighted the lack of economic evaluations that investigate the costs and benefit impacts of innovations in pathology. The remainder of this thesis is dedicated to exploring the cost-effectiveness impacts of digital pathology for breast cancer patients.

Chapter 3: An introduction to digital pathology

The Royal College of Pathology defines Digital Pathology (DP) as *“the use of whole slide imaging to create a digital image that can be used for diagnosis, education and research and will facilitate the development of pathology networks and the introduction of image analysis to assist pathologists in their work”* (RCPATH, 2018a). Digitisation is set to overhaul long established pathology systems and processes. Impacts of its implementation are anticipated to not only be felt in the day-to-day routine practice of the laboratory and in the approaches adopted by pathologists for reporting on tissue specimens, but further beyond these, influencing how clinicians diagnose disease and in turn how patients’ treatment is managed.

When analysing the impacts of DP, a holistic view of the patient pathway needs to be taken to ensure all effects are captured and measured. There are three angles to the logic behind supporting DP adoption; all keeping patient outcomes at the core of their reasoning. Firstly, DP has the potential to accelerate the availability of pathology reports by enhancing efficiency within the laboratory (Griffin and Treanor, 2017), the time to diagnosis and treatment are cut-back and patients are discharged earlier. Secondly, diagnostic accuracy is central to the management of disease; it is essential digitisation of pathology systems does not undermine the current precision levels of a pathologists’ reporting. DP systems should support and where necessary enhance the pathologists’ performance. Computerised image analysis (IA) algorithms have the potential to enhance diagnostic accuracy beyond equivalence; contributing to a quantifiable improvement in accuracy when used together with DP systems that are already in place. The use of DP and IA allows the adoption of a quantitative measure rather than the current subjective approaches to measuring the underlying features and histological characteristics of tissue specimens (Webster and Dunstan, 2014). Clinicians use these results to plan treatment specific to each patients’ individual needs; laying the foundations for better management of disease and improved patient outcomes. Finally, DP can serve as a platform for the development of innovative approaches to the measurement and quantification of disease including the identification of novel biomarkers (Madabhushi, 2009); facilitating the illustration of a comprehensive overview of the individual patients’ underlying disease characteristics, beyond what is currently possible. Clinicians can use this information to further advance the concept of personalised treatment management thus improving patient health outcomes.

While pathology has a significant role to play in the diagnosis, treatment and management of all diseases, the ripple effect of pathology digitisation will be experienced to variable

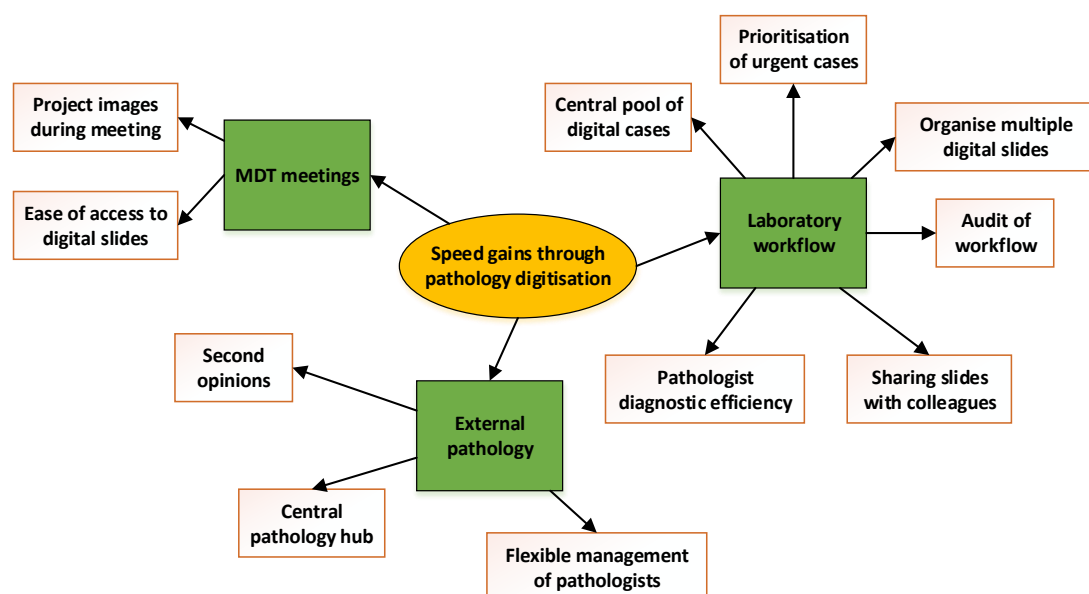
degrees across different disease areas. The scope of the impacts of DP are largely dependent on the inherent shortcomings of each disease pathway, these will differ from one to another. For example; if a specific biomarker is known to be associated with poor diagnostic accuracy on subjective microscope scoring, DP has the potential to contribute to the elimination of errors and improved treatment management for these patients. While for another disease the weakness in the pathway may be related to the speed at which the disease develops, so DP's potential to contribute to laboratory efficiency, and a timelier reporting of results will have the greatest impact on patient outcomes here.

The remainder of this chapter is divided into two sections. The first is dedicated to describing the fundamental features of digitisation that allow it to influence the time to the pathology report being ready. The second section summarises the accuracy implications of digitisation specifically from the breast cancer patients' perspective.

3.1 Speed

Speed can be considered a universal effect of DP, efficient laboratory systems will feed into the tissue management processes of all arriving specimens regardless of the disease area. Differences in the effect size will arise when evaluating the consequent clinical impacts, for some disease areas these are likely to be more pronounced than others. Figure 3-1 provides an overview of the ways by which DP can influence speed in any pathway.

Figure 3-1 Summary of the speed impacts of digital pathology



The effects can be split into 3 groups, through digital pathology's impact on Multi-disciplinary Team (MDT) meetings that take place to discuss patient management, workflow within the

pathology lab and beyond the lab to the mechanisms of interaction with other pathologists and pathology departments. A detailed description of each of these follows.

3.1.1 The laboratory workflow

Currently tissue specimens arrive at pathology labs and glass slides are prepared for analysis by a pathologist using a conventional microscope (CM). The slide preparation phase will be identical for DP but will include an additional step to scan the glass slides prior to the pathologist viewing these on a computer screen rather than under a microscope, as illustrated in figure 3-2. The inclusion of a scanning step in theory is expected to extend the period from the point the glass slide is ready to the point results are reported, in essence delaying rather than speeding pathology processes. Together with the expected increase in cost these time delays will undoubtedly be seen as an obstacle to the adoption of DP; deeming it at outset an unjustifiable technology for investment.

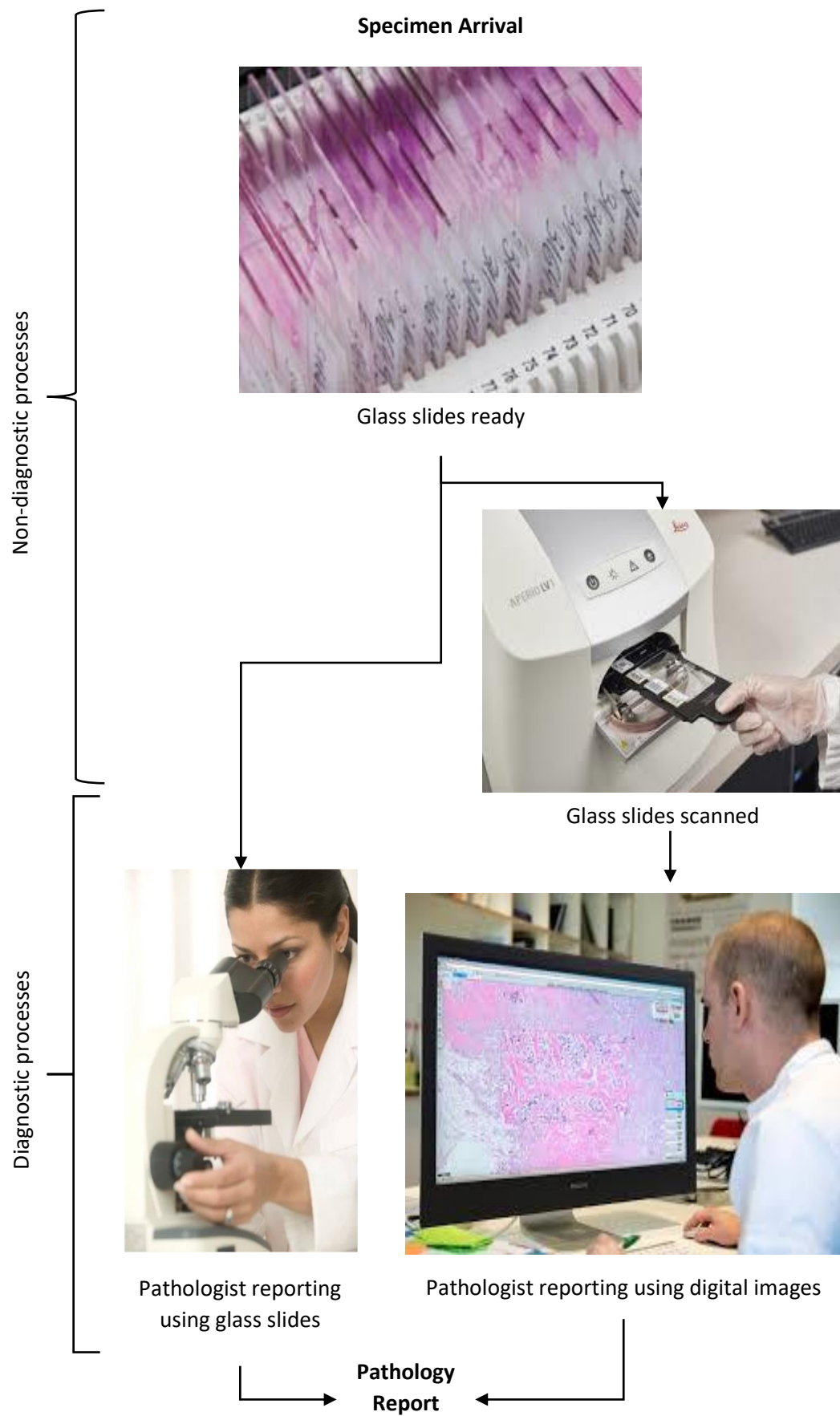
As tissue specimens move through the pathology lab they are subject to both diagnostic and non-diagnostic activities, before the final report with pathology results is available. The non-diagnostic phase encompasses all steps involved in the management and handling of the tissue specimen from the point it arrives at the pathology lab to the point a slide reaches the pathologist for reporting; whether as the original glass slide or in a digital format. The subsequent diagnostic phase covers the steps from glass or digital slide arrival at the pathologist's desk to the pathology report being ready.

3.1.1.1 Non-Diagnostic Phase

Pathologists play a vital role in diagnostic processes; based on their findings treatment decisions are made and implemented. Yet fundamentally, the diagnostic role of the pathologist follows the equally significant non-diagnostic steps that take place prior to the slides reaching their desks. Inclusion of a scanning step will affect the earlier non-diagnostic slide preparation phase of the tissue specimens' pathway. Nonetheless, DP advocates deliberate extensively and are optimistic about the consequences of digitisation on non-diagnostic activities in pathology labs..

This non-diagnostic period will include all the events from specimen receipt to the point of slide arrival with pathologist. These tasks are generally related to a cases' slide preparation, organisation and management, and are chiefly the responsibility of non-pathologist staff. Digitisation will influence this phase in two ways, firstly the processes of preparation and organisation in the steps leading up to arrival at the pathologists' desk for analysis are predicted to run more efficiently and secondly through the introduction of flexibility in the way the pathology workload is allocated when changing over from physical to digital slides.

Figure 3-2 Conventional microscope versus digital pathology processes



Focusing on case organisation and management, the literature describes several potential points in the pathology non-diagnostic workflow that can be advanced with DP. Once scanned, images are immediately available to be viewed at any computer screen in the lab, without these having to be picked up by or dropped off to the pathologist; saving time in the process (Griffin and Treanor, 2017). Most cases will have more than one associated glass slide; these are better organised when digital since they are now easily stored together reducing the possibility of misplacement and labelling errors (Williams et al., 2017). Images can eventually be attached to the corresponding pathology report on the hospital storage systems (Stathonikos et al., 2013). Both implications remove the anxiety and time loss in the laboratory when managing the consequences of loss and breakage since DP eliminates the risk of these occurring. Staff will also have more control over the prioritisation of cases for reporting; when scanning glass slides with DP it is possible to fast track urgent cases using the methods adopted by Thorstenson et al. (2014). They loaded glass slides during the working day onto the scanner in smaller batches so that where necessary urgent cases can be incorporated in each batch leaving a large non-urgent batch to be loaded at the end of the day and scanned through the night ensuring these are ready for pathologists the following morning, thus utilising the non-working hours of the day.

When focused on flexibility in the way the workload is allocated, developments will be seen through the ability of pathologists to access a central pool of digital cases. With CM, a pathologist receives a tray of glass slides for diagnosis (Griffin and Treanor, 2017), in some instances these sit on desks awaiting examination particularly when pathologists are involved with other tasks. This prevents other pathologists who may be free during this time from accessing these same slides and they too will receive a different tray of glass slides for analysis. Digitisation removes the spreading of glass slides across various desks pending pathologist attention (Stathonikos et al., 2013). DP will allow a pathologist who is ready to report to retrieve work from a pool of digital cases moving to a pull rather than a push system of workload allocation. Time is no longer wasted on cases awaiting a free pathologist and are reported on a first come first serve basis. Incidentally, it will be possible to monitor and review the time it takes from slide scanning to report ready, identifying the weak points in the system and so enabling the continual improvement of pathology workflow. Pathologists will also be able to promptly share slides with colleagues especially when seeking a second opinion on ambiguous cases, eliminating the need to move slides between desks and so saving time here also.

Vodovnik (2016)'s study did not formally measure the time differences in non-diagnostic pathology, but suggested based on their observations that they would expect at least a 10% time saving in the overall pathology specimen pathway due to the impact on the non-diagnostic components of the specimen workflow. They equated this time saving to an additional 350 cases per year. Beyond this inference, no further research has been carried out to measure and quantify the difference, specifically in the non-diagnostic pathology time between DP and CM.

Even though the positive impacts of DP on the non-diagnostic steps have not been officially measured, these will undoubtedly, together with the time saving outcomes seen in the diagnostic phase contribute to overall improved efficiency in the laboratory workflow.

3.1.1.2 Diagnostic Phase

Vodovnik (2016) measured the time from slide ready to pathologist report generation and found on average a 6% reduction in this diagnostic period when digital was compared to microscope. This outcome suggests that pathologists will in fact be able to give a diagnosis on a greater number of cases in the equivalent time. Vodovnik (2016) predicted that each pathologist would report on at least an extra 250 cases per year when simply observing the impacts of DP on diagnostic time, as a result enhancing productivity and the overall efficiency of pathology labs. When combined with their projected outcomes for the non-diagnostic phase, in total each pathologist is expected to report on an extra 600 cases per year as a result of digitisation. Stratman et al. (2010) time and motion study further supported these conclusions; showing a 13.4% improvement in the pathologists' diagnostic efficiency with DP apparatus in place. This is approximately an increase of 1.5 cases per day for every full time pathologist (Ho et al., 2014).

The time saving inferences in these studies are attributable to a variety of factors as described by Williams et al. (2017); these include:

- (1) The ability of the pathologist to switch between the various slides of a case instantaneously without having to physically load these onto a microscope
- (2) All activities in the diagnostic process are carried out on a computer screen without having to alternate between this and a microscope
- (3) Slides can now be annotated on screen so any measurements are quicker to carry out
- (4) Pathologists no longer have to adjust the microscope when examining slides, saving time here

- (5) Pathologists are able to immediately retrieve and view older slides digitally especially when dealing with recurrent cases
- (6) Pathologists can access and view molecular test slides concurrently with histology slides

While these appear to be humble modifications; combined they make for an altogether smoother reporting experience for the pathologist and hence bring about the results described by both Vodovnik (2016) and Stratman et al. (2010)'s studies. It is already apparent digitisation can be a means for pathologists to manage a greater workload and potentially reduce the impact of staff shortages. DP can in principal contribute to improving the productivity of the pathologist.

The measured time savings in the diagnostic period are promising for DP but they do not consider the additional scanning steps involved with DP during the non-diagnostic phase. While studies have measured the effect of digitisation on the diagnostic time, no literature was found exploring the overall time changes or the anticipated increase in capacity in pathology labs when taking into consideration the impact of DP on both diagnostic and non-diagnostic activities. No formal analysis has compared the extra time needed for scanning with the time saved due to the overall improved workflow in pathology labs. Nonetheless based on the presented evidence it is safe to conclude that the time saving will outweigh the time gains seen with digitisation. On any given day, a pathologist will have a greater number of available slides to examine as well as have the capacity to analyse and diagnose a greater number of cases. Consequently, a larger number of patients are ready for discussion at the next available Multi-Disciplinary Team (MDT) meeting and so a care plan can be agreed sooner influencing patient outcomes.

3.1.2 MDT meetings

MDTs have brought many benefits to patient care due to clinician collaboration on diagnosis reducing the number of errors (Kane and Luz, 2013), as a result there is a growing demand for these meetings. Hospitals are frequently setting up new MDTs as well as increasing the frequency of those existing (Kane et al., 2007), consequently clinicians are dedicating a greater amount of their limited time in attendance and preparation for these as well as reporting on a larger number of cases. The digitisation of pathology systems is reported to influence MDT meetings both directly and indirectly. All improvements in the pathology lab workflow discussed previously are likely to have a knock on effect on this part of the patients care pathway. Since evidence shows an increase in lab productivity this inevitably leads to a greater number of cases being ready and available at MDT for discussion sooner.

Simultaneously, instances where patients are due to be discussed but pathology reports are not ready are largely diminished. Clinicians save time in picking up a case only for it to be deferred to the following meeting due to missing pathology data. These are indirect methods by which digitisation will influence MDT. An increase in the availability of cases to be discussed due to digitisation in pathology will further burden meetings, potentially causing a backlog especially when they are already at full capacity.

Some of the ways in which to ensure MDT meetings run with greater efficiency in order to be able to accommodate the increased work load resulting from pathology digitisation and make better use of clinicians' time will come as a direct consequence of digitisation itself. A survey by the National Cancer Action Team (Taylor and Ramirez, 2009) found less than 75% of MDTs had the technology in place to project images from glass slides. This no longer becomes an issue if slides are already on a digital platform since it is safe to assume all hospitals will have the apparatus in place to project images from a computer screen. The survey of MDT members also found over 98% of participants strongly agreed that they needed access to retrospective images and pathology reports (Taylor and Ramirez, 2009). With DP, all images and reports associated with the patient can be retrieved and viewed instantly; decreasing the number of MDT deferrals to give time for laboratory staff to locate the required information. The prompt accessibility of pathology images is of great benefit particularly when discussing recurrent cases and images related to a previous episode need to be brought up. DP will allow clinicians to view all slides for the case concurrently without loading slides individually onto a microscope (Stathonikos et al., 2013); an even greater benefit to teams without access to microscope projecting facilities who would normally group around a microscope to view more than one glass slide. Prior to the meeting glass slides are normally located and picked up by a staff member for cases due to be discussed at MDT, this takes less time with digitisation since physical slides are not collected and only digital images organised (Williams et al., 2017). All MDT members are able to view annotations made by the pathologist of clinically interesting findings they would like to highlight (Thorstenson et al., 2014) since these can now be made directly onto the slides with DP, supplementing the information available for treatment management.

While it is difficult to quantify the direct influence on productivity with each projected use value to MDT, there is proof to support meetings running with greater efficiency with DP apparatus in place. A study exploring the experience of DP introduction at a Swedish hospital (Thorstenson et al., 2014) provided evidence MDT meetings were managed with greater effectiveness resulting in a 66% increase in productivity. Each case was discussed in less time

and so in the 90 min allocated time for the meeting, 25 cases had agreements in place for next steps of action rather than the usual 15 cases complete in the same time slot.

3.1.3 Reporting beyond the laboratory

There are several circumstances under which pathologists have to send slides away for external reporting. A case may be difficult to interpret and so a second opinion is sought or the opinion of subspecialty pathologist is needed but not available on site and so the case needs to be sent away for assessment (Têtu and Evans, 2013). Currently glass slides are delivered at additional cost and time, with the risk of damage along the way. The same applies when pathology labs do not have the required facilities to carry out certain molecular tests and so the glass slides must be sent away for reporting. With DP, scanned slides can be sent electronically reducing the time it takes to get an external opinion from days to hours (Al-Janabi et al., 2012) thus allowing for the swifter reporting of results while preserving the glass slides.

Digitisation will allow a more malleable approach to work load allocation as described earlier, this is further augmented by the potential flexible management of pathologists. Offsite reporting with DP can be a way in which to overcome current staff shortages. Without the need for a microscope pathologists no longer have to be at a desk to report on a case but are able to examine and give a diagnosis from anywhere and at any time as long as a computer screen with the appropriate software is available, since glass slides are no longer needed and digital images are sufficient for diagnosis (Maras, 2015). This opens the door for more flexible working practices allowing pathologists to work from home or work part time or from different locations (Ho et al., 2014). DP will enable outsourcing of excessive work due to staff shortages since it is now possible to achieve a diagnosis within a reasonable amount of time (Thorstenson et al., 2014). This has already to an extent been seen in Sweden where Kalmar hospital scan all their slides citing the need to work with other labs owing to the country wide shortage of pathologists (Pantanowitz et al., 2011). In future, the creation of central pathology hubs incorporating pathologists across all specialities will be the epitome of efficiency in pathology practice. Hospital labs are able to scan and send slides here for reporting ensuring the improved consolidation and management of workload across all hospitals in a trust.

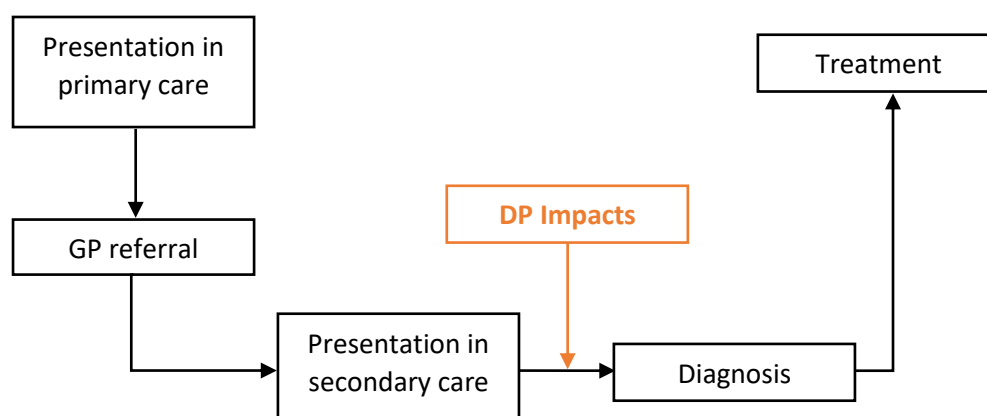
No study explored the effect of flexible pathology workforce management on productivity but a search of the literature found one investigation exploring DPs impact on external reporting time. Vodovnik (2015) compared distance reporting with digital to onsite reporting using a microscope; they found DP reduced the average reporting time to 3.13 days

compared to the current 3.90 days with a microscope onsite. The proportion of cases reported in 3 days or less increased from 40.56% to 72.25% with external digital reporting. Pathologists can deliver a diagnosis sooner using DP remotely than when using a CM onsite. This study did not compare current distance reporting processes using CM and so did not take into account the current activities involved in the delivery of slides which include packing, delivery offsite, unpacking and delivery to second pathologist (Sectra, 2015). Preferably Vodovnik (2015) should have compared distance reporting using DP with current distance reporting practices; undoubtedly an even greater time saving would have been observed.

3.1.4 The patient experience

A patient with suspected cancer progresses through a series of appointments and tests in both primary and secondary care as illustrated in figure 3-3, before true disease status is confirmed and the appropriate treatment interventions are made.

Figure 3-3 Overview of the patient cancer pathway



Minimising the intervals between each point in the pathway guaranteeing diagnosis and treatment initiation as soon as possible is of immense importance both to the patient and to the NHS. UK cancer survival rates fall behind many other countries with delays in diagnosis and treatment as one of the causes for this discrepancy (Foot and Harrison, 2011). Even what may seem to be humble decreases in any unnecessarily prolonged period of the cancer pathway could result in cancer diagnosis treatment initiation before metastatic growth (Roope and AFOM). Evidence for time savings during both the diagnostic and non-diagnostic phases of the specimens' pathology pathway, efficiency savings in MDT and time savings when reporting beyond the lab have all been reported with digitisation. These factors affect the time period between presentation in secondary care and diagnosis. It is unlikely DP will influence any other phase of the patient pathway. Secondary care delays up to point of diagnosis fluctuate between cancer type and vary dependent on source of referral (Allgar

and Neal, 2005). GP referred cases experience greater delays in this period than those cases presenting through screening and colorectal cancers take nearly twice as long on average than suspected breast cancers to reach a diagnosis (Allgar and Neal, 2005). Digitisation can be an avenue through which to explore levelling this period across all cancer types regardless of the point of referral. With DP patients will benefit through having their pathology reports available earlier, thus a case can be discussed at MDT with a care plan agreed and treatment initiated sooner.

Waiting for results can be an anxious period for many patients; a systematic review focusing on anxiety during the diagnostic phase of suspected breast cancer found overwhelming evidence to support this (Montgomery and McCrone, 2010). Any attempts to shorten this period will have positive impacts on patient wellbeing especially for those cases confirmed to be benign on diagnosis; patients can return to normal life no longer worrying about the prospects of undergoing further cancer treatment. As well as anxiety, prolonged periods at this point can be an inconvenience for patients especially if not managed appropriately. It is recommended patients are scheduled for outpatient appointments (OPA) on the same or following day of the MDT meeting where their case is discussed (National Cancer Action Team, 2010). It is expected that all required reports including those from pathology will be available with the MDT to plan the course of action, enabling a clinician to communicate this to the patient during the upcoming OPA. Unfortunately, some cases are not discussed at MDT due to instances of missing pathology information. These cases are inevitably postponed to the following meeting yet the patient is still due to come in for their scheduled OPA. In this scenario, patients are merely informed their results are not complete and so their OPA is rescheduled. This wastes both clinicians and patients' time and enhances anxiety already experienced while awaiting confirmation of their disease status. For the patient, DP potentially signifies enhanced quality of life through reduced anxiety, inconveniences avoided and most importantly better survival outcomes.

3.2 Accuracy

Diagnostic accuracy measures express the nearness of the pathology result to its true value (Raab and Grzybicki, 2010), interpretations of pathology slides resulting in a diagnosis that does not reflect the underlying disease state (Raab et al., 2005) can influence a patients' treatment therapy. Improvements in diagnostic reporting accuracy lead to either prescribing treatment that previously would have been missed or avoiding treatment that would otherwise have been given incorrectly.

Unlike the speed implications of DP previously discussed which feed into all sorts of disease pathways, accuracy implications are specific to disease areas. Each disease area will have a unique set of histological features and disease characteristics that need to be measured for clinicians to make a diagnosis and to plan treatment. Dependent on the current reported accuracy with CM for each of these, DP systems have the potential to at least provide non-inferior results by measuring diagnostic concordance and where possible enhance the pathologist's reporting for those particular categories with recognised deficiencies by measuring diagnostic accuracy. These effects can only be quantified independently by disease area due to the breadth and scope of the influence pathology has within medicine. The remainder of this chapter will review the available research evidence for accuracy implications of DP from the perspective of the breast cancer patient.

Numerous results are included in the pathology reports of breast cancer patients; pathologists grade the cancer, measure the tumour size and margins, determine lymph node (LN) involvement and score both the ER and HER2 biomarkers. Only when all this information is available can clinicians plan and design the treatment package appropriate for the patient. Khazai et al. (2015)'s study retrospectively reviewed 1970 breast cancer cases referred to their cancer centre for second review during 2010. They found 11.47% of all cases to have significant discrepancies between the original report results and those in the second review, corresponding to 225 patients diagnosed wrongly or given incorrect treatment. These discrepancies will undoubtedly affect patient care and outcomes. The quantitative rather than subjective approach to scoring taken with DP can in theory contribute to the reduction of inter-observer variability and enhance the precision of the pathology results reported.

Numerous validation studies have compared the diagnostic accuracy of pathologists when examining and reporting on tissue specimens using digital slides to when this is done directly on the original glass slide using a microscope. Mills et al. (2017)'s evaluation examined 510 surgical pathology cases across five different organs; accuracy levels were similar across both methods. Snead et al. (2016)'s validation study also proved digital accuracy to be non-inferior to microscope when they analysed the results for 3017 cases reported on by pathologists using both approaches. A systematic review by Goacher et al. (2016) reported a mean diagnostic concordance rate of 92.4% based on 38 studies comparing CM and DP that met its inclusion criteria.

There are likewise several studies comparing CM and DP specifically focused on breast cases. A recent randomised study compared the accuracy of breast pathologist's reporting on glass

slides to the digital set-up (Elmore et al., 2017), using a consensus reference standard. The study authors found the accuracy of the results observed using glass slides to be 3-5% superior, undermining the adoption of DP systems. However, they do acknowledge the pathologists' lack of experience with using DP as a possible explanation for this discrepancy. A validation study that also focused on breast pathology reported 98.8% clinical concordance when comparing CM and DP (Williams et al., 2018a). All pathologists that were involved in studying the breast cases were initially provided with one-to-one training and were required to view and discuss a training set of twenty challenging cases before they proceeded onto live diagnoses using the digital platform (Williams et al., 2018a). Three further studies compared CM to DP diagnoses on breast pathology cases. Al-Janabi et al. (2012)'s study reported a 93% concordance level between both approaches. Of the discrepant cases, only 1% would have clinical implications for the patients and their outcomes. Campbell et al. (2014) reported an overall concordance rate of 97% when comparing reporting on breast cases using DP to using CM. Finally, Reyes et al. (2014) included three pathologists, each examining the same slides under a microscope and on a screen. Their concordance rates were 99%, 99% and 96%.

The evidence so far supports the non-inferiority of DP when compared to CM for diagnosing breast cases. However most studies did not focus on the accuracy implications of the individual histological features that are scored to make the overall diagnosis. These need to be measured for both conventional and digital slides independently and compared to a reference standard to highlight where DP will make the most difference to the pathology reports of breast cancer patients and thus their diagnosis and treatment.

Figure 3-4 Summary of the areas of potential accuracy gains with DP

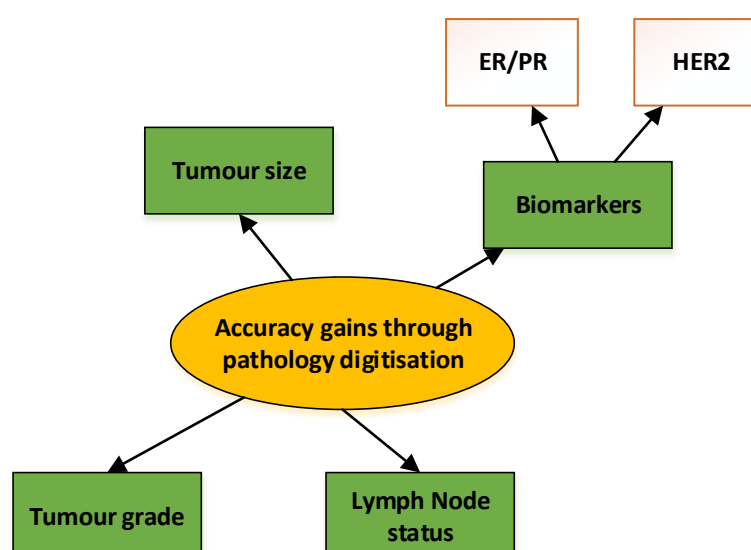


Figure 3-4 provides an overview of the potential accuracy gains reported in the literature for breast cancer patients when digital pathology systems are in place. They centre on improvements in tumour grading, determination of lymph node status, scoring molecular biomarkers and measuring tumour size. The remainder of this section summarises the evidence for each of these.

3.2.1 Biomarker profile

There are three tissue based biomarkers that are routinely measured and analysed in order to plan breast cancer treatment, these are oestrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor 2 (HER2) (Ulaner et al., 2016).

3.2.1.1 ER/PR

The overexpression of ER/PR will guide hormonal treatment in breast cancer patients. Patients with low levels of these biomarkers will not benefit from hormonal therapy whereas patients with raised levels will respond to treatment. It is essential that these are scored accurately to avoid cases of over or under treatment and thus ensure the best possible outcomes for breast cancer patients.

UK-NEQAS determined the frequency of ER/PR positivity by analysing data submitted by 200 of its UK laboratories. According to this audit, in the UK 73.42% of breast cancer cases are ER positive and 58.1% of breast cancer cases are PR positive (Rhodes et al., 2000a), however high levels of variability in positivity rates exist between laboratories (Rhodes et al., 2000b). Hammond et al. (2010) set out a series of practice guidelines and recommendations to overcome these inconsistencies; these included the endorsement of the use of quantitative IA techniques for scoring ER/PR.

Testing for ER/PR is carried out using immunohistochemistry (IHC) staining, considered the gold standard in scoring for breast cancer cases (Yaziji et al., 2008). In essence, this identical staining method is used across both approaches and the difference lies around the medium the pathologist uses to examine these identical slides. However with DP, pathologists have the opportunity to use IA algorithms to support their reporting, allowing for a more quantitative approach to biomarker scoring (Madabhushi and Lee, 2016).

Nassar et al. (2011b) analysed inter-observer variability for six pathologists across two sites, each scored ER/PR using both conventional and digital IA. They found reproducibility of results with IA superior to that achieved with current microscopic examination. Lloyd et al. (2010) reported 100% equivalence in their study comparing pathologist's subjective ER

scoring to results attained using quantitative IA. Turbin et al. (2008)'s study also presented evidence for consistent concordance between IA and microscopes for scoring ER.

Many of the studies evaluating the impacts of digitisation on ER/PR results do not calculate diagnostic accuracy parameters such as sensitivity and specificity but rather examine concordance between and reproducibility levels across both approaches. Measuring the accuracy of IHC scoring for the ER/PR biomarkers, whether with CM or DP is challenging due to the lack of test that can be considered the reference standard against which comparisons can be made. Using pathologist's subjective scoring as the reference standard in assessing the analytical validity of the quantitative scoring of newer IA algorithms is problematic (Aeffner et al., 2017), this assumes subjective scoring to represent the height of accuracy. Suggestions have been made for the use of pathologists' consensus score or a combination of pathologist and IA scoring as the reference standard, since together they are expected to be superior to each alone, thus combining the strengths of both methods (Aeffner et al., 2017).

While there is perhaps evidence to support equivalent or enhanced precision with digital IA, it is not so obvious if there are advances in accuracy.

3.2.1.2 HER2

Herceptin treatment is recommended as part of the care plan for HER2 positive breast cancer patients, whereas negative cases will not respond to this treatment. Herceptin is an expensive drug with an associated risk of cardio-toxicity, so it is essential results from pathology labs are accurate enough to guide treatment, limiting side effects to the patient and costs to the NHS. The guidelines for HER2 testing in breast cancer recommend the use of IHC staining as the first step in determining HER2 status, with FISH used where results are inconclusive (Rakha et al., 2014). Unlike ER/PR testing, where no test exists that is thought of as the reference standard, for HER2 testing FISH has been correlated with superior accuracy and is more likely to generate reproducible results when compared to IHC (Ross, 2011), for these reasons it is generally considered the gold standard in HER2 testing.

DP systems have the potential to bring IHC HER2 results closer to those of FISH HER2, ensuring a greater number of breast cancer patients are prescribed Herceptin in line with their true HER2 status. Several studies have demonstrated greater accuracy of IHC when pathologists use IA with DP rather than conventional microscopes (Holten-Rossing et al. (2015), Dobson et al. (2010) and Skaland et al. (2008).

3.2.2 Lymph node status

At some point in their pathway, all breast cancer patients will have their LNs examined for evidence of metastasis. If a palpable mass is detected in the axilla at the initial screening stage, a LN biopsy is taken alongside that from the breast. Otherwise, all patients with confirmed breast cancer in their CB will undergo a sentinel lymph node biopsy (SLN) with surgery to establish evidence of LN involvement. The SLN is where cancers cells are first likely to spread from the breast (Heerdt, 2018) and the presence of cancer cells here is an important prognostic indicator for breast cancer patients (Lyman et al., 2005).

The SLN biopsy replaced the more invasive axillary lymph node dissection (ALND) procedure, fewer LNs are removed for analysis with evidence of enhanced identification of metastasis in SLN compared to ALND due to the focused nature of pathological examination (Giuliano et al., 1995). The accuracy of the SLN biopsy has been extensively researched and proven comparable to the gold standard ALND procedure. Miltenburg et al. (1999)'s meta-analysis showed a 98% concordance rate between SLN biopsy and ALND and a 5% false negative rate across all included studies. Pesek et al. (2012) much larger meta-analysis focused on studies only including SLN biopsies with a negative result, they estimated a 7% (6.1 to 7.9%) false negative rate using a random effects model for meta-analysis.

The examination of fewer LNs with SLN biopsies promoted the adoption of more advanced techniques for the detection of cancer cells than those used with ALND, this made for a more intensive and time consuming pathological examination process (Maaskant-Braat et al., 2011) increasing the work load on already strained pathology labs. Digital pathology systems can be used as a platform for the development of approaches to minimise the challenges presenting to pathology labs when examining and reporting on SLN biopsies (Holten-Rossing et al., 2017).

International researchers were asked to develop computer algorithms for the detection of metastases in SLN biopsies of breast cancer patients as part of the Cancer Metastases in Lymph Node challenge 2016 (CAMELYON16) (Bejnordi et al., 2017). 32 algorithms were submitted, the area under the receiver operating curve (AUC) for metastases classification ranged from 0.556 to 0.994 with four algorithms having AUCs that were statistically different to the AUC of pathologist scoring (Bejnordi et al., 2017), showing some evidence of increased accuracy with computer algorithms. Holten-Rossing et al. (2017) also developed a computer algorithm for SLN analysis and compared this to standard manual pathologist examination. They reported 100% sensitivity and 68.9% specificity i.e. no false negative cases but a risk of over treatment due to the false positive cases. However, their sensitivity and specificity

values were calculated using conventional pathologist scoring as the reference standard so once again this is analysis of concordance rather than superior accuracy. Holten-Rossing et al. (2017) propose the adoption of their algorithm as a screening tool to separate out the negative from the positive cases and use manual assessment on all positives cases which will include the false and the true positives. This alone will eliminate 58.2% of the work involved in SLN biopsy analysis, Holten-Rossing et al. (2017) measured and compared the time spent on SLN examination using microscopy and using their algorithm. The average time was reduced from 6.88 to 2.88 minutes. Due to the already advanced accuracy of SLN evaluation, the greater benefit of digital systems to measuring this feature of breast cancers may come with the minimisation of the time-consuming analysis process.

3.2.3 Tumour grade

Breast cancers are divided into three groups when assigning their grade; grade 1 cancers look most similar to normal cells and are slow growing, grade 3 look most unlike normal cells and are fast growing and in between are grade 2 breast cancers (Breast Cancer Care, 2018). Oncologists consider the breast cancer grade when planning a patient's neoadjuvant and adjuvant treatment; especially when deciding on the administration of chemotherapy. However, there are concerns around the lack of reproducibility when grading breast tumours (Elston and Ellis, 1991) due to the subjective techniques used by pathologists when assigning a grade (Rakha et al., 2018b).

Boiesen et al. (2000) examined the reproducibility of grade results across seven pathology laboratories in Sweden; concordance ranged from 57% to 77% providing evidence for only moderate levels of reproducibility between laboratories. Longacre et al. (2006) reviewed pathologist's scoring for each of the individual breast cancer features, including grade. They showed evidence of both inter-observer variability; ranging from 12.2% to 38.6% and intra-observer variability ranging from 0% to 13%. The highest level of disagreements in both analyses corresponded to grade 2 tumours. Longacre et al. (2006) also analysed the accuracy of the grade classification by comparing the study pathologists' scoring to that of the study chair, which they regarded as the reference standard. The following results were reported on the mean and range of accuracy for each grade:

- Grade 1: 83.3% (75 to 100%)
- Grade 2: 64.6% (50 to 83.3%)
- Grade 3: 92.3% (79 to 100%)

The highest levels of discrepancy are once more associated with the scoring of grade 2 tumours. Approximately 17% of cases scored as grade 1 could in fact be grade 2 and thus are potentially missing out on chemotherapy treatment. Over 35% of grade 2 tumours are scored incorrectly, some of these are likely to be grade 1 and so not considered for chemotherapy treatment, while others are expected to be grade 3 and would be routinely given chemotherapy treatment. Approximately 8% of grade 3 tumours could in reality be grade 2, these patients are not routinely prescribed chemotherapy but their grade is considered alongside other characteristics of the breast cancer when deciding on the treatment plan.

Rakha et al. (2018a) compared the grading of breast cancers when using DP to when pathologists use CM techniques. The results reported in their study are adapted to give the percentage agreement levels in table 3-1.

Table 3-1 Agreement between DP and CM when grading breast cancers

Digital	Microscope		
	Grade 1	Grade 2	Grade 3
Grade 1	85%	31%	3%
Grade 2	14%	62%	30%
Grade 3	1%	7%	67%

Discrepancies in grading breast cancers exist between CM and DP, with the highest levels of discordance around grade 2 cancers. However, the authors have not measured the accuracy of each method by

comparing the results to a 'gold standard'. From these results alone, it is not possible to ascertain whether DP represents a more superior approach to grading or is in essence equivalent or maybe even inferior. Rakha et al. (2018b) analysed intra-observer variability when pathologists grade each of the digital images a second time after a three-month interval. They reported overall a 77% level of agreement when grading breast cancers; specifically 77% for grade 1, 70% for grade 2 and 86% for grade 3 cases. These fall below the intra-observer agreement rates described by Longacre et al. (2006) when grading breast cancers using conventional techniques, suggesting lower levels of precision with DP compared to CM. Nonetheless, it is difficult to make robust conclusions regarding the non-inferiority of DP to grade breast cancers when very limited primary research is available on which decisions can be made.

Beyond digitisation, computer algorithms that quantify the underlying scores that together makeup the assigned grade are anticipated to become available (Madabhushi, 2009). No study so far has compared and reported on the grading of breast cancers using computer-aided quantification to conventional subjective grading.

3.2.4 Tumour size

The size of the breast tumour influences the patient's treatment plan. It is measured using imaging techniques such as mammography or ultrasound prior to surgery. Together with clinical examination, clinicians will use the tumour size to decide whether a patient undergoes a mastectomy or a wide local excision. DP is not likely to influence the surgery recommended but it can play a role in the assessment of tumour margins post-surgery and thus the decision for further re-excision. IA algorithms can be used alongside DP systems to differentiate between invasive and uninvolved cells within the excision margins (Cruz-Roa et al., 2017). While the theory underpinning the use of algorithms with DP is understood, the primary research comparing the results of margin analysis using CM with results from digital techniques and the consequent impact on re-section rates is limited. A significant fundamental issue in tumour margin analysis is the lack of consensus on the definition of a positive margin (Emmadi and Wiley, 2012) and while there are guidelines set by the Association of Breast Surgery there is considerable variation in current practice across hospitals in the UK and Ireland (Tang et al., 2017). Before the DP impacts on margin assessment and the consequent re-excision rate can be determined, breast pathologists need to be clear on the underlying thresholds. Validation studies for the measurement of this feature of breast pathology reporting are non-existent and further primary research is needed before any conclusive decisions can be made on the value of DP with or without IA when examining breast margins.

3.3 Conclusion

Throughout the current literature it is clear the generation of digital images of glass slides alone will not bring about accuracy improvements when measuring the features of breast cancer that are required to make decisions around treatment and disease management. DP slides are identical to CM slides but they are in principal viewed using different mediums; it is only with supplementary technology such as quantitative image analysis and computerised scoring techniques that DP can really begin to make a difference to the accuracy of results and allow the development of novel diagnostic approaches. Digitisation alone has its greatest impact on workflow and laboratory systems, as long as DP systems are proven to have equivalent diagnostic concordance without undermining current diagnostic accuracy when compared to conventional microscopic techniques, there is an argument to support the adoption of digital systems based on efficiency implications alone.

Knowledge of the individual effects seen with digital pathology are of limited value unless the clinical impacts to the patient are known. The time and accuracy consequences of

digitisation need to be measured by quantifying the implications they will have for the patient, whether digitisation influences their time to diagnosis or time to treatment or changes patient care plans as a result of changes to diagnoses. Further research and evaluation is needed to give value to the digitisation of pathology through the perspective of breast cancer patients by determining the clinical utility of this pathology innovation.

Chapter 4: The breast cancer pathway

4.1 Introduction

Chapter 3 highlighted the ways by which the digitisation of glass slides can influence the examination and the analysis of breast cancer tissue specimens in the pathology lab. Whether through its accuracy or time implications; these effects must be measured and evaluated whilst keeping the breast cancer patients at the centre of all decision making. The differences in pathology results observed when moving from CM to DP are only of true value if there is a concomitant measured benefit for the patient. Likewise, efficiency in pathology labs as a result of digitisation should be assessed in relation to their effects on the patient. Outputs of pathology labs are not directly linked to patients in the same way, as for example treatment interventions; pathology results and their availability will guide surgery and systemic treatment plans, which will in turn influence patient outcomes.

In order to be able to identify, measure and give value to the effects of digitisation for the patient, it is important to first clearly understand the breast cancer pathway and the underlying relationships between pathology and the different events that take place within the patient's course of management in secondary care. An appreciation of this network of activities will highlight how modifications and innovations in pathology through digitisation will influence patient outcomes. Further to this, mapping the pathway will identify any shortcomings in the system that could perhaps be overcome through advances in pathology; setting the scene for further research of novel technologies to overcome these shortfalls.

4.1.1 Information sources

There are two sources of information that can be drawn from to clearly illustrate the breast cancer pathway. Firstly, evidence-based guidelines such as those published by the National Institute for Health and Care Excellence (NICE) and secondly, the experiences of clinicians in the management of patients under their care. Ideally, clinicians should comply with published recommendations so that clinical practice reflects guidelines and all patients are offered identical opportunities in the diagnosis, management and treatment of their disease, creating an environment of high quality care across all NHS services. In reality this is not always the case and disparities exist in the management of breast cancer patients in the UK (Purushotham et al., 2001).

Guidelines represent tools that can be used by NHS services to ensure consistency in clinical practice and management of disease. They are *“systematically developed statements to assist practitioner and patient decisions about appropriate healthcare for specific clinical*

circumstances” (Lohr and Field, 1990). They are developed following a meticulous process of activities that involve the identification of the best available research evidence through systematic review of the literature, consultation with experts and piloting of guidelines to establish the feasibility of adhering to them in practice (Thomas, 1999). Guidelines are regularly reviewed and updated to reflect the most up to date findings; bridging the gap between research and practice.

While numerous guidelines are available in healthcare for different disease areas including breast cancer, avoidable variations in clinical practice are still evident indicating inconsistencies in the implementation of evidence-based recommendations. Irregularities in care and survival outcomes for breast cancer patients continue to exist across England (All-Party Parliamentary Group, 2018). This is in spite of previous recognition of this concern and the proposal of general health service recommendations in the ‘Five Year Forward View’ report by NHS England (2014a) as well as recommendations specific to cancer in the ‘Achieving World Class Cancer Outcomes’ report by the Independent Cancer Taskforce (2015) to overcome variation in care. Development of guidelines alone is not sufficient, systems need to be in place to ensure these are disseminated and implemented appropriately (Kennedy et al., 2010).

The key motivation behind the development of evidence based guidelines is the enhancement of quality of care to patients as well as the maximisation of patient health outcomes (Woolf et al., 1999). There is evidence that patients managed in accordance with guidelines score better on outcomes than those managed based on clinician’s experience alone. Sacerdote et al. (2013) recognised the presence of improved quality of care for breast cancer patients post implementation of clinical practice guidelines. Furthermore, Andreano et al. (2017) demonstrated an increased probability of 5year survival for breast cancer patients whose management followed an evidence based care pathway compared to those that did not.

The impacts of interventions in pathology can only be evaluated when there is a clear understanding of the relationship between pathology services and other activities in the breast cancer care pathway. It is important to first conceptualise the breast cancer pathway structure to appreciate the role pathology has to play within the grander scheme of breast cancer patient care. Following this, a network of the chain of events starting with pathology digitisation as the generic process can be illustrated to depict the transformation of outputs from here into target processes, followed by clinical processes and finally patient outcomes.

The measurement of patient outcomes must be done within the framework of a pathway that is based on both guidelines and practice to ensure precision in the evaluations of pathology interventions. When interventions are evaluated within an evidence based pathway that may not represent clinical care, this may lead to misleading results especially where inconsistencies exist between both. A breast cancer pathway modelled on both guidance and practice represents a more realistic depiction of care, based on which pathology impacts can be evaluated. Besides an understanding of both versions of the breast cancer pathway will identify and highlight the differences between both. Gaps as well as barriers to the application of guidance in practice are recognised and recommendations made to remove these.

4.2 Aims

There are four aims to this chapter:

1. To develop an understanding of the breast cancer pathway that is based on both guidelines and clinical practice at University Hospitals Coventry and Warwickshire (UHCW)
2. To identify the discrepancies in clinicians' adherence to published recommendations and where possible provide explanations and appreciate the challenges for implementation
3. To visually summarise the activities in the breast cancer pathway and the decisions that control the interaction between them to be used as the foundation for evaluations of any interventions along the pathway
4. To depict the network of the chain of events that connect pathology digitisation as an intervention to patient outcomes

4.3 Methods

There were two stages to the development of the breast cancer pathway. Firstly, UK guidelines and recommendations were used to develop a pathway that reflects evidence-based guidelines. Secondly, the expert opinion of the clinicians at UHCW involved in the management and treatment of breast cancer patients was sought. Two breast cancer care pathways were developed so that comparisons could be made between them. Each pathway represents the course of activities and the underlying decisions made from the point a patient with suspected breast cancer is referred to secondary care to the point of treatment planning and initiation.

4.3.1 Methods for the descriptions of the evidence-based pathway

An online literature search was carried out to identify all available UK guidelines that are designed to be used by clinicians during the diagnosis, treatment and management of breast cancer cases. A list of key words was compiled and used in the website searches:

Breast AND (cancer or HER2 or ER or oestrogen or grade or chemotherapy or radiology or pathology or surgery or lymph nodes or Herceptin or invasive or in situ or malignant or early or primary)

In particular, searches were carried out for NICE guidance, pathways and technology appraisals as well as Public Health England (PHE) professional guidance for their NHS Breast Screening Programme (NHSBSP). In addition, searches for breast cancer related guidance was carried out in the websites of the Royal College of Pathologists, Royal College of Surgeons and the Royal College of Radiologists. The websites of the various charities that provide information and support to breast cancer patients were also searched. These included Breast Cancer Care, Macmillan Cancer Support, Breast Cancer UK and Cancer Research UK. Any material circulated by these organisations is based on research findings and UK recommendations. A list of all the publications identified are listed in table 4-1. A large degree of overlap was expected between the sources listed, where disparate recommendations are made in more than one publication but related to the same step in breast cancer management, information from the most recent will be incorporated into the pathway.

All publications were reviewed and information from these extracted to inform the evidence-based breast cancer pathway. This pathway models the epitome of care quality that NHS trusts are expected to adopt in order to maximise patient health outcomes.

Table 4-1 List of publications to inform the evidence-based breast cancer pathway

Source	Publication	Reference
NICE	Breast Cancer (Quality Standard 12)	(NICE, 2011a)
	Early and locally advanced breast cancer: adjuvant therapy	(NICE, 2014)
	Early and locally advanced breast cancer: diagnosis and treatment (Clinical Guidance 80)	(NICE, 2009)
	Early and locally advanced breast cancer overview	(NICE, 2011b)
	Hormonal therapies for the adjuvant treatment of early oestrogen receptor positive breast cancer (Technology Appraisal Guidance 112)	(NICE, 2006a)
	Improving outcomes in breast cancer (Cancer Service Guideline 1)	(NICE, 2002)
	Trastuzumab for the adjuvant treatment of early-stage HER2-positive breast cancer (Technology Appraisal Guidance 107)	(NICE, 2006b)
PHE	NHS Breast Screening Programme: Consolidated standards	(PHE, 2017)
	Quality assurance guidelines for surgeons in breast cancer screening	(NHS BSP, 2009)
	Clinical guidance for breast cancer screening assessment	(Borrelli et al., 2016)
	Quality assurance guidelines for breast cancer screening radiology	(NHS BSP, 2011a)
	Quality assurance guidelines for breast pathology services	(NHS BSP, 2011b)
	Reporting, recording and auditing B5 core biopsies with normal/benign surgery	(NHS BSP, 2007)
Royal Colleges	Appendix D : TNM classification of tumours of the breast (UICC TNM 8)	(RCPATH, 2018b)

	Pathology reporting of breast disease in surgical excision specimens incorporating the dataset for histological reporting of breast cancer	(Ellis et al., 2016)
	Guidelines for non-operative diagnostic procedures and reporting in breast cancer screening	(Lee et al., 2016)
	Best practice diagnostic guidelines for patients presenting with breast symptoms	(Willett et al., 2010)
	Surgical guidelines for the management of breast cancer	(ABS, 2009)
	Postoperative radiotherapy for breast cancer: UK consensus statements	(RCR, 2016)
Patient Information	Breast Cancer Care: Understanding your pathology results	(Breast Cancer Care, 2013)
	Breast Cancer Care: Primary breast cancer	(Breast Cancer Care, 2016)
	Understanding breast cancer	(Macmillan, 2016)
	Breast Cancer	(Cancer Research UK, 2017)
Other	Updated UK recommendations for HER2 assessment in breast cancer	(Rakha et al., 2014)

4.3.2 Methods for the description of the clinical-practice pathway

From the evidence-based pathway, it was possible to identify the medical specialities within secondary care that breast cancer patients are exposed to. Contact with both a consultant and trainee pathologist at UHCW had already been established early on in the study, details for the remaining clinicians that are responsible for breast cases at UHCW were requested from the pathologists. The clinicians were contacted for interview to seek their expert opinion on breast cancer care. The following clinicians were interviewed to inform the clinical-practice pathway:

- Breast care nurse (surgical)
- Breast care nurse (oncological)
- Consultant breast pathologist
- Consultant breast radiologist

- Consultant breast surgeon
- Consultant pathologist
- Medical oncologist
- MDT lead clinician
- Trainee pathologist

Semi-structured interviews were carried out with each of the clinicians listed; these comprised general questions around breast cancer care in addition to speciality specific questions about the section of the pathway relevant to each clinicians' training area. Clinicians were given the opportunity to highlight inconsistencies between guidelines and their practice. If any conflicts were identified between guidelines and clinical practice, rationalisations for their presence were explored with the clinician through further questioning. Where there was ambiguity around steps in the evidence-based pathway, perhaps due to absent guidance, these points were also discussed.

All interviews were recorded and transcribed. As each member of the breast cancer care team was interviewed, the evidence-based pathway was updated to reflect the practice of clinicians at UHCW. On completion of interviews, the clinical practice breast pathway describes current breast cancer care at a large teaching hospital.

In summary two pathways were depicted in Microsoft Visio, the first representing the breast cancer pathway based on guidelines alone and the second on clinical practice.

4.3.3 Methods for the description of the network of digital pathology processes

Chapter 3 summarised the published literature exploring the evidence of the impacts of pathology digitisation for breast cancer patients, this information was merged with the breast cancer pathway that is described later in this chapter.

The two sources of information were used to design a process pathway portraying digital pathology and the processes within the breast cancer pathway it is expected to connect to in order to indirectly influence patient outcomes. The resultant pathway is a chain of events starting with digital pathology as a generic process and ending with appropriate patient outcomes. From the breast cancer pathway, the targeted and clinical processes that sit between were identified.

4.4 Results

4.4.1 Evidence-based pathway

The published guidelines and recommendations for breast cancer management were reviewed; these were used to inform the evidence-based pathway. The following section describes this. Based on guidelines the patient pathway through secondary care can be broken down into several phases as outlined in figure 4-1. Guidelines acknowledge that patients with suspected breast cancer can be referred for assessment in secondary care via two sources, either through their general practitioner if they present in practice with signs or symptoms of breast disease or they can be asymptomatic cases that are referred through routine breast screening. First patients should be assessed to determine the nature of their disease; this should be followed by a discussion of the case at MDT where management plans should be agreed. There are two types of breast cancers, ductal carcinoma in situ (DCIS), where cancer cells have not spread beyond the lining of the ducts to surrounding tissue and invasive breast cancer (IBC), where cancer cells have spread to the surrounding tissue (Cancer Research UK, 2017). For DCIS cases, the recommendations for management include surgical intervention only and for IBC they involve a combination of both surgery and systemic therapy. The guidelines were explicit in defining that each breast cancer type should be treated differently. The phases in figure 4-1 can each be further broken down for a more granular understanding of the published recommendations that fall under each of the assessment and treatment phases of the breast cancer pathway.

4.4.1.1 Triple assessment

Assessment of suspected breast cancer cases referred to secondary care for investigation should be carried out in the breast clinic during a single hospital visit. Figure 4-2 summarises the three recommended interventions that together make-up the assessment carried out to verify breast cancer status. At this stage of the breast pathway all suspected cases are expected to be managed in the same way regardless of their source of referral, since it will not yet be known if a tumour is IBC or DCIS or whether it is even cancerous.

Published guidelines divide triple assessment between clinical examination, imaging and biopsy. In the first step, the clinician should examine the breast and axilla of the patient for any obvious palpable lumps. This should be followed by imaging of the breast to identify the location and provide initial indications of the size and multi-focality of the breast tumour. Where discrepancy exists between clinical assessment and imaging, guidelines recommend an MRI scan is carried out in addition to the initial imaging procedure. If there is obvious evidence of malignancy on clinical examination due to the presence of a palpable lump in

the axilla, the patient is also recommended to undergo an ultrasound of the axilla. The final recommended step in triple assessment involves the removal of a tissue sample from the breast by biopsy and only if there is evidence of malignancy on ultrasound of the axilla, a biopsy should be taken from the lymph nodes. Decisions at MDT are made based on the results of this assessment process.

Figure 4-1 Steps in suspected breast cancer management based on published guidelines

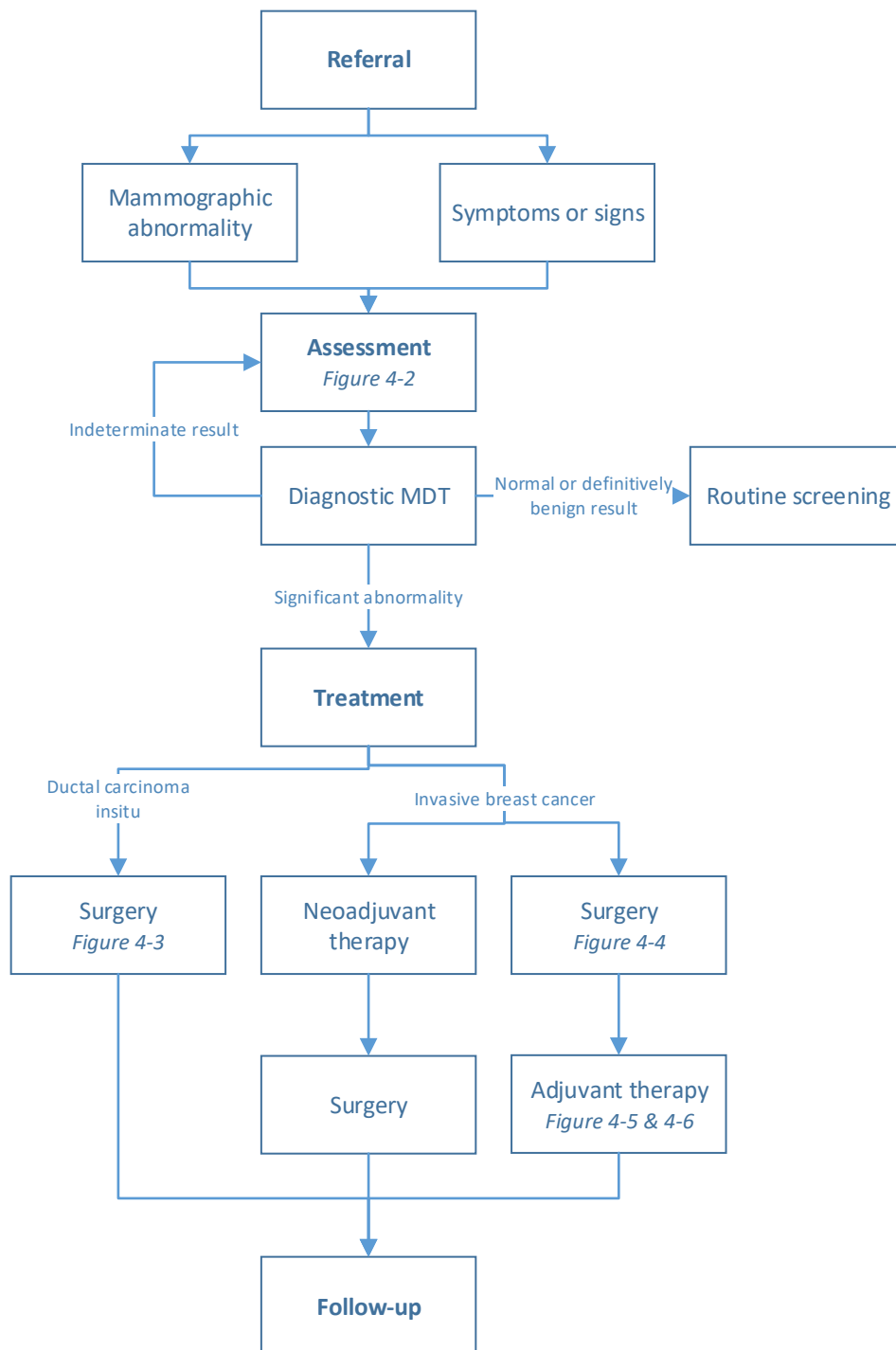
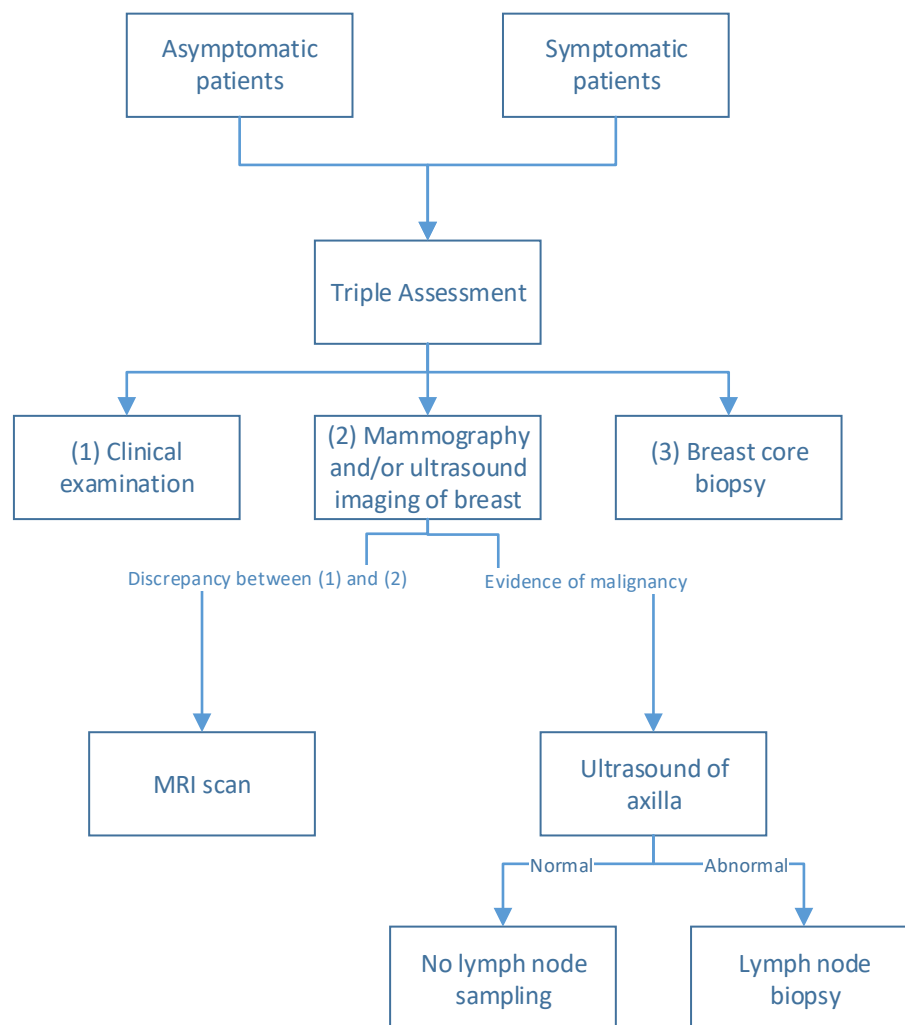


Figure 4-2 Guidelines for triple assessment



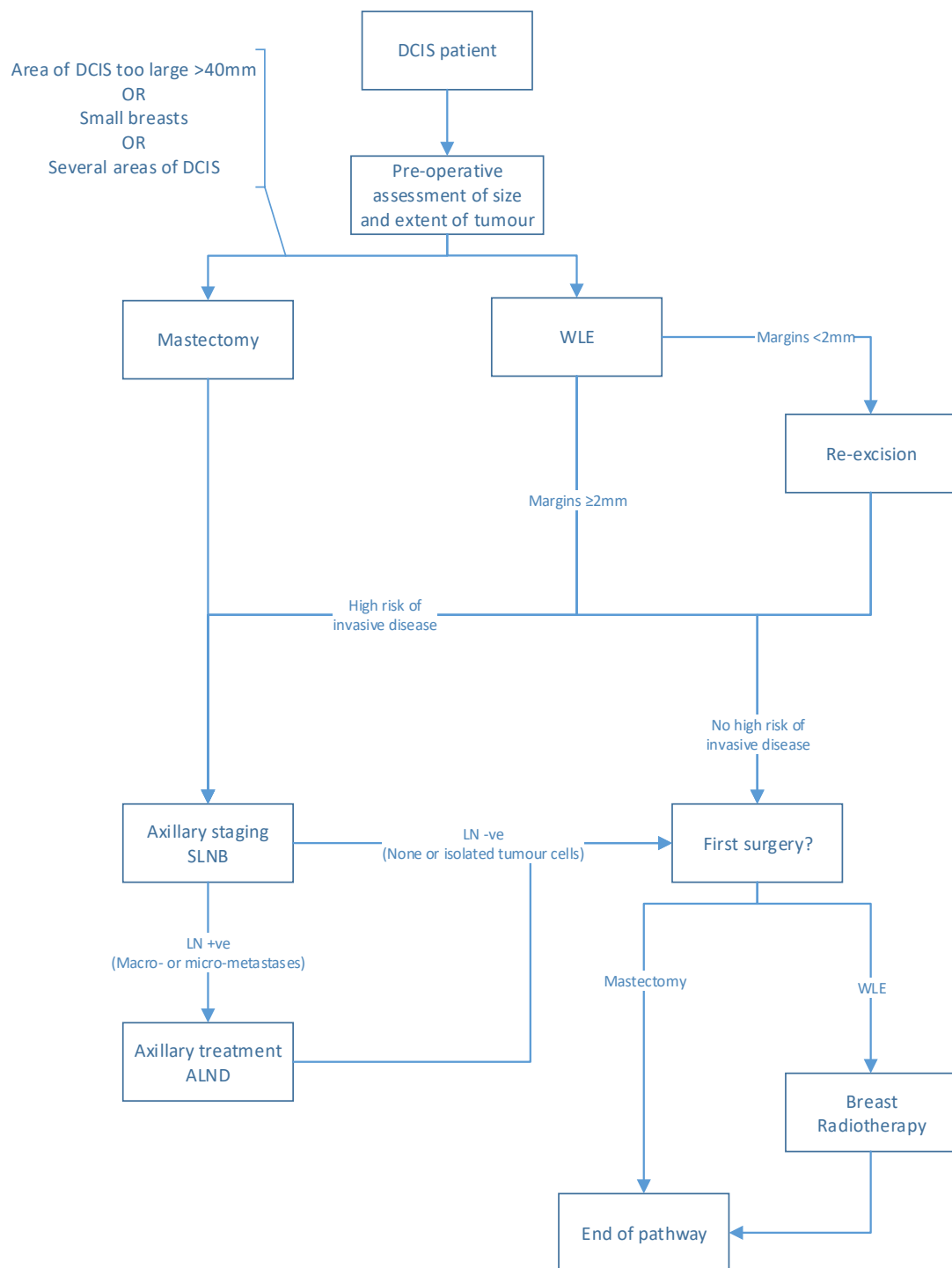
4.4.1.2 Breast cancer treatment

The breast cancer by this stage has been confirmed as either DCIS or IBC and the guidelines recommend the commencement of treatment as appropriate. Guidelines for the management of invasive and DCIS cases are reported separately. DCIS should be considered for surgical and radiotherapy interventions as part of their treatment plan whereas IBC cases are recommended a combination of surgical, radiotherapy and systemic interventions dependent on various pathology results.

4.4.1.3 Surgery and radiotherapy

For DCIS and IBC tumours, surgery is the first-line recommendation followed by other therapies where appropriate. Radiotherapy and surgery are closely linked; the decision to offer radiotherapy is largely associated with the surgical intervention undergone by the patient. Figure 4-3 outlines the surgical and radiotherapy guidelines for DCIS cases and figure 4-4 **Error! Reference source not found.** and figure 4-5 do the same for IBC cases.

Figure 4-3 Overview of surgery and radiotherapy guidelines for DCIS cases



The surgery recommended for DCIS cases is largely dependent on the size and extent of the tumour, mastectomies should be carried out where the ratio of tumour to breast is large otherwise patients should undergo wide local excisions (WLE). On pathological analysis of the WLE surgical specimen, a clear margin of at least 2mm is recommended in NICE guidelines, if less than this the patient should return to surgery for re-excision until a clear margin of greater than or equal to 2mm is achieved.

The guidelines do not endorse axillary staging as routine practice for DCIS patients unless they undergo a mastectomy or at high risk of invasive disease where a WLE is recommended. The axillary staging should take place alongside their main surgical interventions. High risk DCIS patients are defined as patients with a palpable lump in the axilla on clinical examination (NICE, 2009). As outlined in figure 4-3, only DCIS patients undergoing a WLE procedure are recommended radiotherapy post-surgical intervention.

The surgical and radiotherapy recommendations for IBC cases are discussed in greater detail in the guidelines than for DCIS cases, this is largely due to the underlying nature of this breast cancer type. IBC have already spread into the breast tissue when compared to DCIS cases, they are treated with greater urgency and as a result these patients are exposed to a greater number of tests and examinations and have a more complex treatment and management pathway.

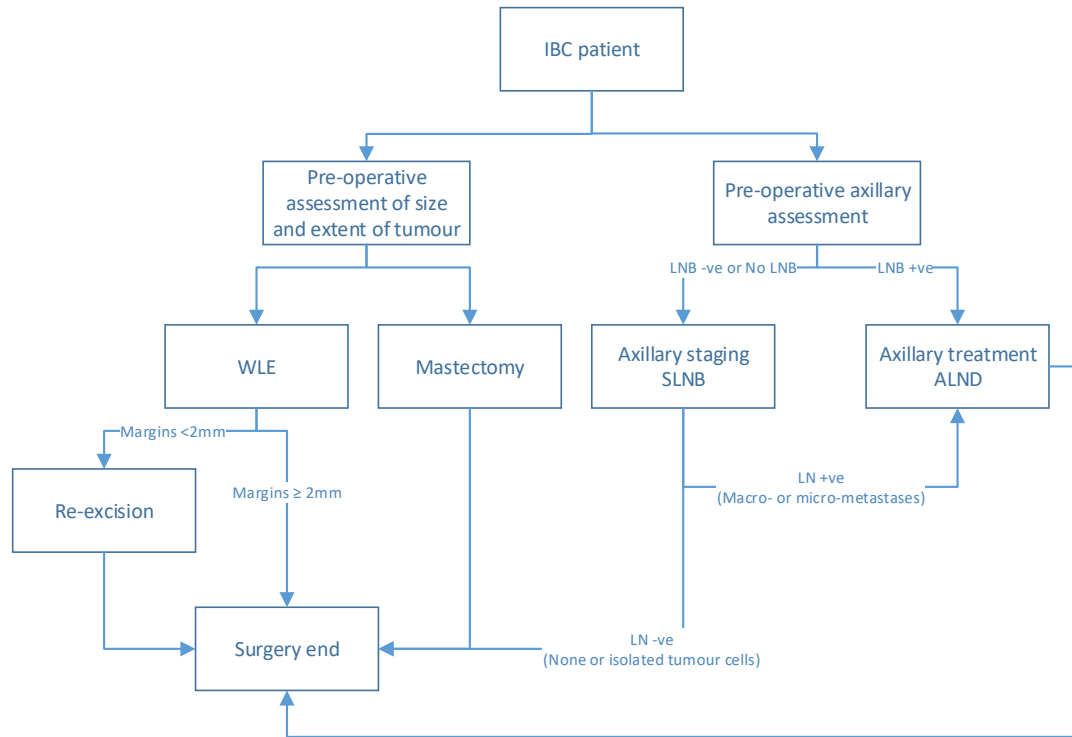
Similarly to DCIS cases, the choice between mastectomy and WLE is based on the size and extent of the tumour but no further guidelines are given. Throughout the literature, the terms 'informed decision making' are used but no threshold on the size of the tumour are explicitly stated. It is clear the choice of surgery should be made based on MDT experience and informed discussion with the patient. Where WLE has been carried out, consistent with DCIS cases, the surgical specimen margins are measured and a cut off of 2mm is recommended to determine the need for further re-excision.

Unlike DCIS cases, axillary staging is recommended routine practice for IBC cases. All patients should undergo an axillary surgical intervention alongside their breast surgery, dependent on pre-operative assessment and results. Patients with a previous positive lymph node biopsy are accordingly recommended an ALND, whereas those with a negative lymph node biopsy or no previous lymph node sampling should be considered for a SLN biopsy. Where LNs are positive on pathological examination of the SLN, an ALND is recommended as a second surgical intervention. The surgical guidelines for IBC patients are summarised in figure 4-4. The radiotherapy published recommendations are also extensive for IBC patients as summarised in figure 4-5.

Treatment with radiotherapy is dependent on both the breast surgery undertaken and the nodal status of the patient. The guidelines recommend breast radiotherapy following all WLEs whereas patients undergoing a mastectomy are only considered for radiotherapy subject to their risk of recurrence. Low risk cases are not recommended radiotherapy at all and high risk cases are recommended chest wall but not breast radiotherapy, in between

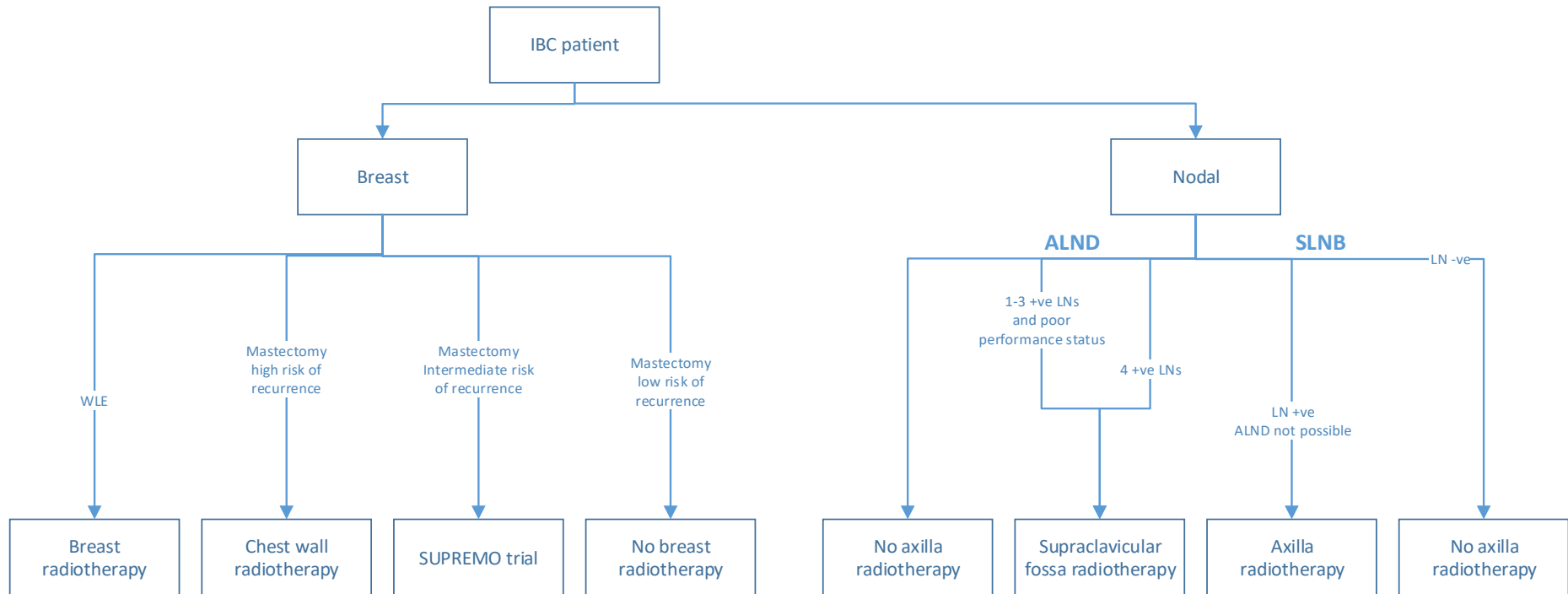
patients with intermediate risk are not normally recommended radiotherapy intervention but the SUPREMO trial investigates the use of chest wall radiotherapy in this group of patients (Kunkler et al., 2008) and so they should be offered the choice to take part in the trial.

Figure 4-4 Summary of the recommended surgical pathway for IBC cases



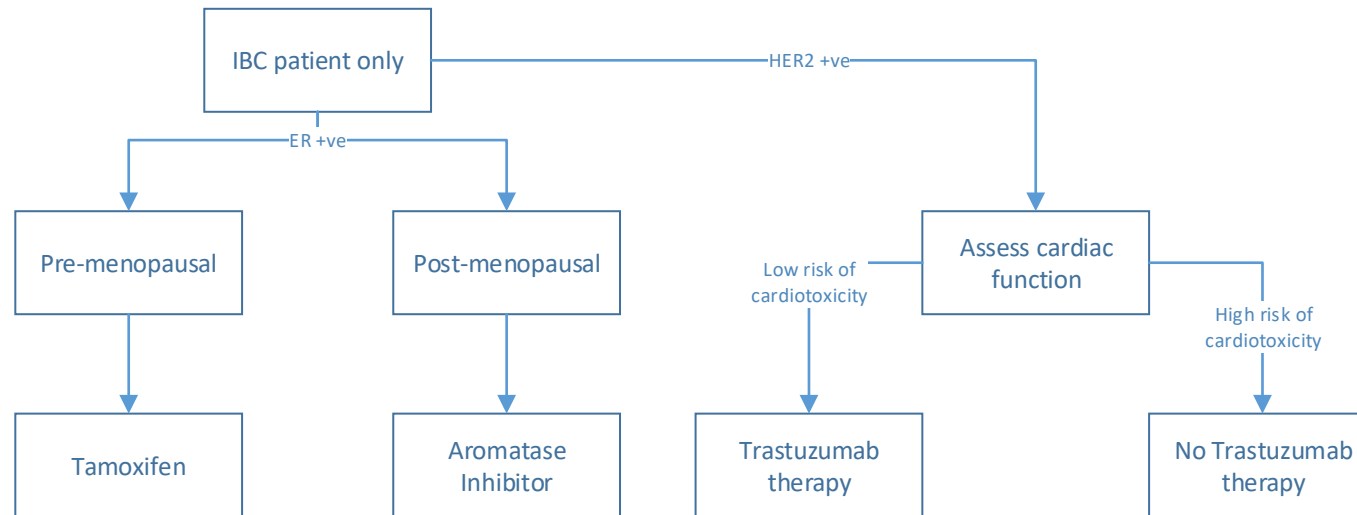
Radiotherapy to the axilla is mentioned only when LNs are positive on SLN biopsy but an ALND is not possible. Post ALND, radiotherapy to the axilla is not considered but radiotherapy to the supraclavicular fossa is recommended dependent on the number of positive lymph nodes after pathological examination of the ALND.

Figure 4-5 Overview of the radiotherapy guidelines for IBC cases



4.4.1.4 Systemic therapy

Figure 4-6 ER and HER2 therapy treatment guidelines



Systemic interventions should be offered to IBC patients only; these are a combination of hormone therapy, Trastuzumab (Herceptin) and chemotherapy. The decision to prescribe the first two is relatively straightforward as outlined in figure 4-6. However, recommendations on the chemotherapy regimens to be given to IBC patients was not as clear-cut to learn from the guidelines alone.

Chemotherapy

Chemotherapy is prescribed by oncologists to reduce the risk of recurrence and improve overall survival in women that are being treated for breast cancer (Early Breast Cancer Trialists' Collaborative Group, 1998). Whilst Herceptin and hormone therapy are included in breast cancer treatment plans in response to the result of a given biomarker test, the decision to administer chemotherapy is not as prescriptive as it is for the other systemic therapies. Chemotherapy is clearly recommended and discussed throughout the published guidelines ((Breast Cancer Care, 2016), (NICE, 2014), (NICE, 2011a), (ABS, 2009), (NICE, 2009) and (NICE, 2002)) however the circumstances under which chemotherapy should be prescribed and the regimens and dosages to be given were not so clear.

The most concise summary of the types of patients that should be recommended chemotherapy came from Breast Cancer Care (2016)'s patient information leaflet on primary breast cancer. Breast Cancer care is a specialist UK charity providing information and support to those affected by breast cancer. In the leaflet they list the features of a patient's breast cancer that are most likely to predispose them to chemotherapy treatment:

- Tumour larger than 2cm
- Grade 3 IBC
- LN positive
- HER2 positive

Further to this, NICE (2002)'s *Improving outcomes in breast cancer* publication recommends chemotherapy for all women assessed as being at intermediate or high risk of recurrence. The risk of recurrence is closely related to the presence of the cancer features listed in Breast Cancer Care's information leaflet.

Three publications NICE (2014), NICE (2011a) and Breast Cancer Care (2016) reference the use of Oncotype DX, a tool that predicts the likelihood of cancer recurrence (Carlson et al., 2013) in patients with intermediate risk, defined as those with ER positive, HER2 negative and LN negative breast cancers (NICE, 2011a). Results of this assay are used to support chemotherapy decisions. Overall, it can be deciphered from the published guidelines that breast cancer patients with low risk of recurrence should not be routinely prescribed chemotherapy since this group is not mentioned in the chemotherapy sections of guidelines. Patients with intermediate risk may or may not be prescribed chemotherapy and the guidelines recommend the use of prediction tools such as Oncotype DX to support clinician decision making. Patients with high risk of recurrence are likely to be prescribed

chemotherapy. Beyond these relations between risk of recurrence and chemotherapy treatment a few other recommendations can also be taken from the guidelines. These are summarised in the table below.

Table 4-2 Summary of indications for chemotherapy in breast cancer

Hormone status	
Both ER positive and negative breast cancers can be considered for chemotherapy. If status is positive and hormone therapy is given clinicians need to measure whether there is any benefit to chemotherapy especially in the following cases: <ul style="list-style-type: none"> • Patient has a low risk of recurrence • ER expression is high • Other existent causes of mortality e.g. old age 	(ABS, 2009)
HER2 status	
Chemotherapy given alongside Herceptin treatment if HER2 status is positive.	(Breast Cancer Care, 2016)
Triple negative	
Triple negative breast cancers i.e. HER2, ER and PR negative can be treated with chemotherapy	(Breast Cancer Care, 2016)
LN status	
Patients with LN positive breast cancers should be offered docetaxel Patients with LN positive breast cancers should not be offered Paclitaxel	(NICE, 2009)
Patients with LN positive breast cancers can be given a combined regimen of docetaxel, doxorubicin and cyclophosphamide Patients with LN positive breast cancers are not recommended paclitaxel	(NICE, 2014)
Risk of recurrence	
4-8 cycles of chemotherapy are recommended for women with intermediate or high risk of recurrence.	(NICE, 2002)

There are no further chemotherapy recommendations in the reviewed guidelines beyond those listed in table 4-2. It is not possible to put together a chemotherapy decision algorithm giving details of the regimen and dosages to be prescribed and under which circumstances. Chemotherapy treatment decisions appear to be complex; clinicians must consider overall breast disease rather than simply one biomarker suggesting that clinician experience plays a

key role in treatment planning rather than the use of clearly defined chemotherapy treatment decision algorithms.

Neoadjuvant treatment

The final consideration in breast cancer treatment is the use of neoadjuvant therapy, the administration of systemic therapy prior to surgery rather than the usual vice versa methods. The guidelines are ambiguous about the circumstances under which such an approach would be recommended, there is no clear indication in any of the publications. Only NICE (2014) mention patient preference suggesting that choices around the organisation of breast cancer treatment are left to the clinicians and discussion with their patients.

However, it is clear that the underlying motivation behind this format is the evasion of a mastectomy procedure. Initial systemic therapy is expected to reduce tumour size, allowing the patient to undergo a WLE rather than a mastectomy on surgery. The same combination of drugs that would be used in the adjuvant setting would be recommended in the neoadjuvant (NICE, 2002).

4.4.2 Clinical practice pathway

The evidence-based pathway described so far was further modified to reflect clinical practice at UHCW, a large teaching hospital. The foremost clinicians involved in breast cancer management and treatment were identified, contacted and interviewed and their interviews transcribed. As each clinician was questioned, the pathway was updated.

It was possible with the information gathered to develop a more detailed sense of the pathway from patient arrival in secondary care through to treatment initiation and management. To some extent, some clarity was gained around the elements of the evidence-based pathway with associated ambiguity, particularly the neoadjuvant and chemotherapy sections. In general the findings taken from the clinician interviews supplemented the evidence based pathway and allowed a deeper understanding of the breast cancer patient experience.

Figure 4-7 provides an illustrative overview of the clinical practice pathway, the basic frame of this is very similar to the illustrative summary of the evidence-based pathway given in figure 4-1. However, through clinician discussion it was possible to precisely identify the points of pathological examination and assessment as well as to deduce the frequency and positions of both surgical and oncological patient consultation to really understand how the activities in the breast cancer pathway are interconnected.

Figure 4-7 Overview of the clinical practice breast cancer pathway

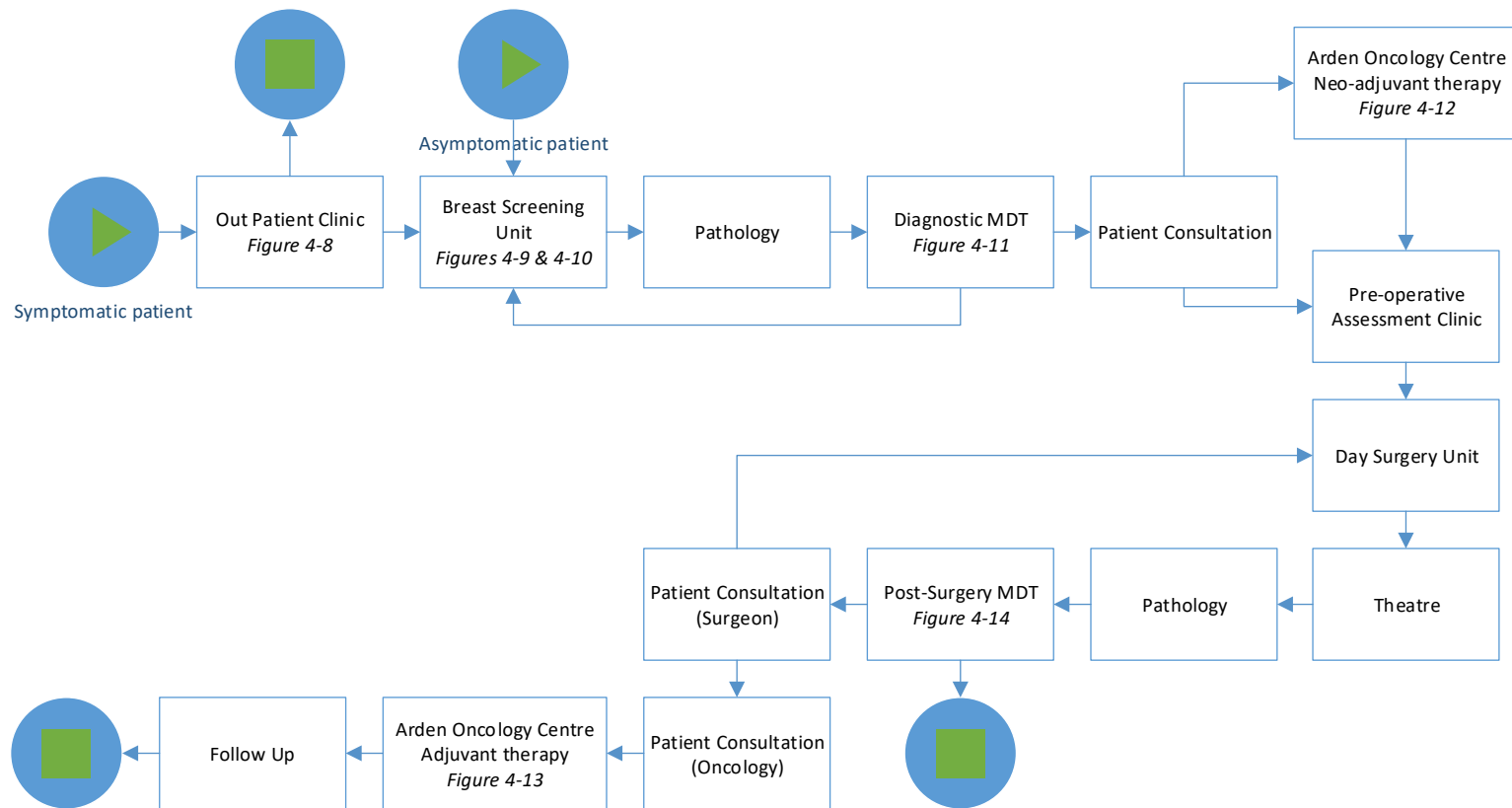


Figure 4-8 Outpatient clinic (symptomatic patients)

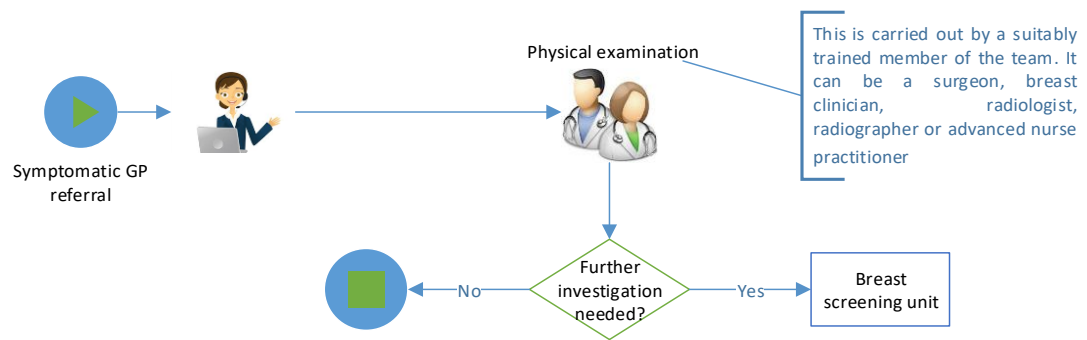


Figure 4-9 Breast screening unit (symptomatic patients)

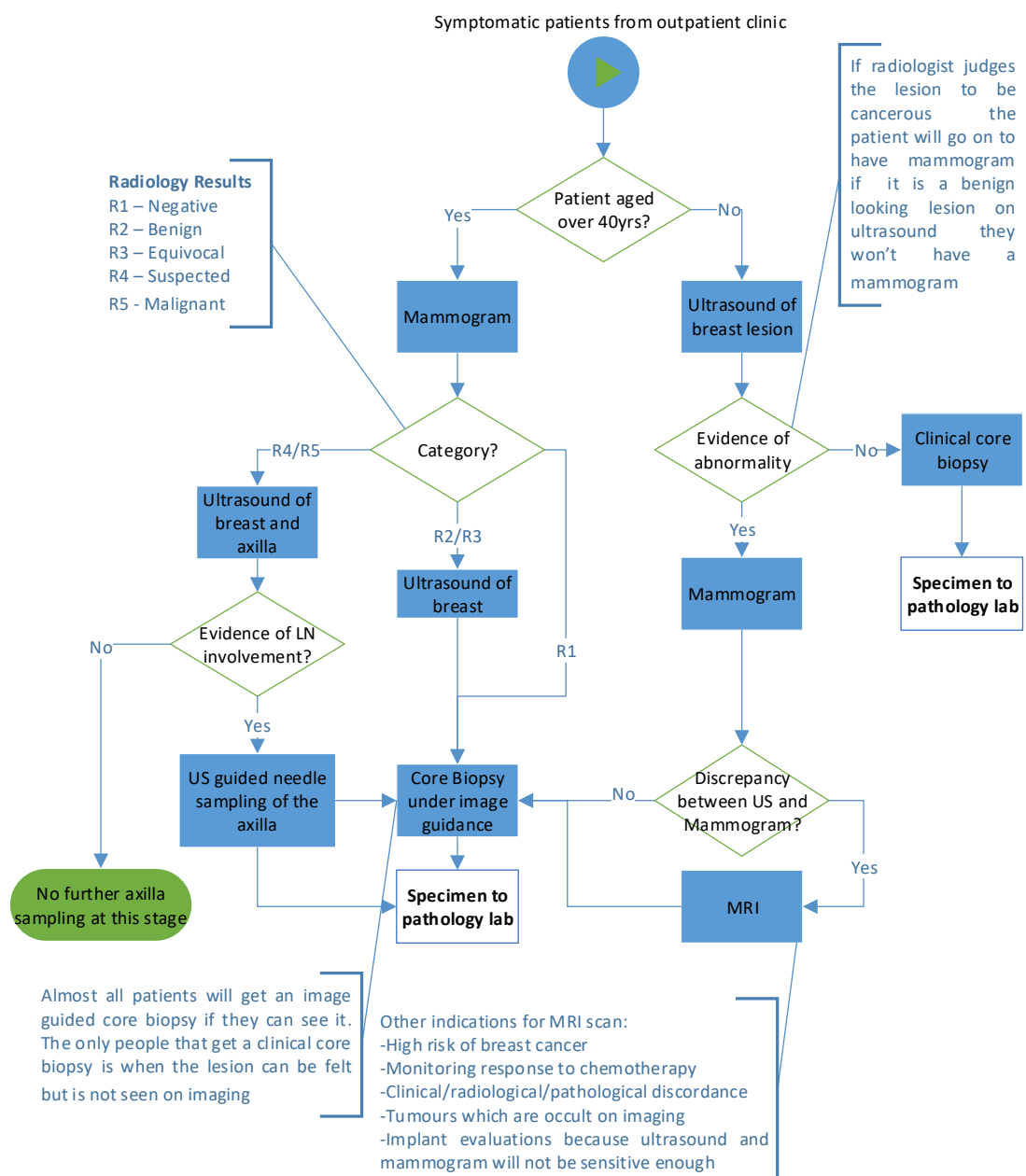
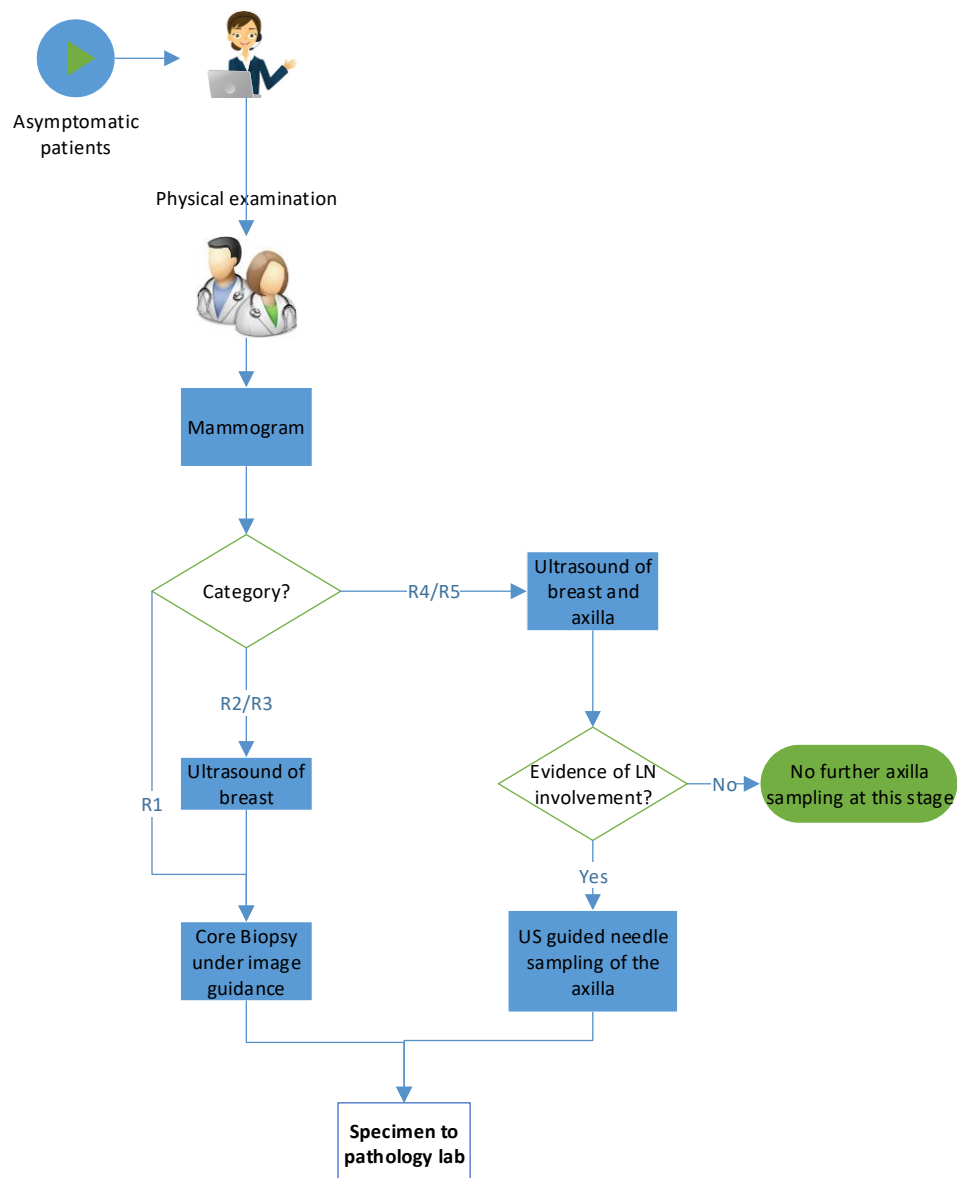


Figure 4-10 Breast screening unit (asymptomatic patients)



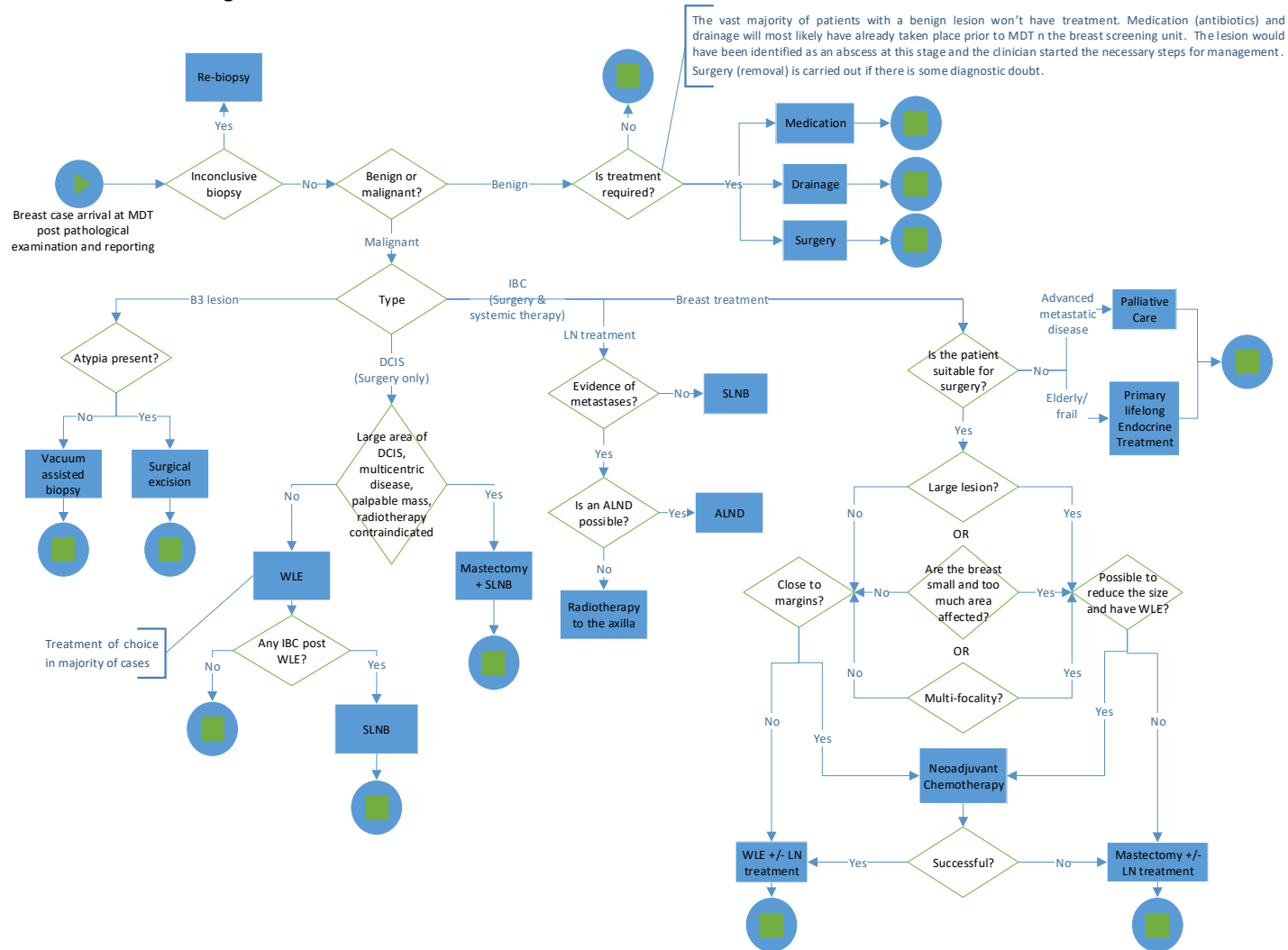
Figures 4-8, 4-9 and 4-10 summarise the steps in the assessment of symptomatic and asymptomatic breast cases on their arrival in secondary care. Both groups undergo triple assessment, encompassing physical examination, imaging and biopsy, consistent to what is already learnt by reviewing the guidelines. Physical examination is a standard procedure carried out in the outpatient clinic for symptomatic patients and in the breast-screening unit for asymptomatic patients. Asymptomatic and symptomatic patients over 40 years follow identical pathways through the breast-screening unit, post physical examination. The guidelines stipulate mammography and/or ultrasound of the breast as the imaging modalities of choice. Conversations with clinicians clarified the circumstances under which each or both would be recommended. Asymptomatic patients and symptomatic patients over 40 years will undergo a mammogram as their first-line imaging procedure, the breast

lesion category reported at this stage will dictate subsequent imaging. Negative lesions will undergo no further imaging, benign and equivocal lesions will have an ultrasound of the breast only and suspected or malignant lesions will have an ultrasound of both the breast and axilla. Post imaging all breast cases, from normal to malignant lesions will have a tissue sample removed through core biopsy of the breast in addition to biopsy of the lymph nodes where there is evidence of involvement on the ultrasound image. Symptomatic patients under 40 years follow a slightly different pathway through the breast-screening unit. Their initial imaging procedure is an ultrasound of the breast with a mammogram undertaken if the radiologist suspects breast cancer on ultrasound. This will be followed by a core biopsy.

In essence, there were no contradictions between the evidence base recommendations outlined earlier and clinical practice, the guidelines permitted a general depiction of triple assessment which was supplemented by clinician interviews to provide a more comprehensive understanding of the underlying activities and decisions that are involved in this section of the breast pathway.

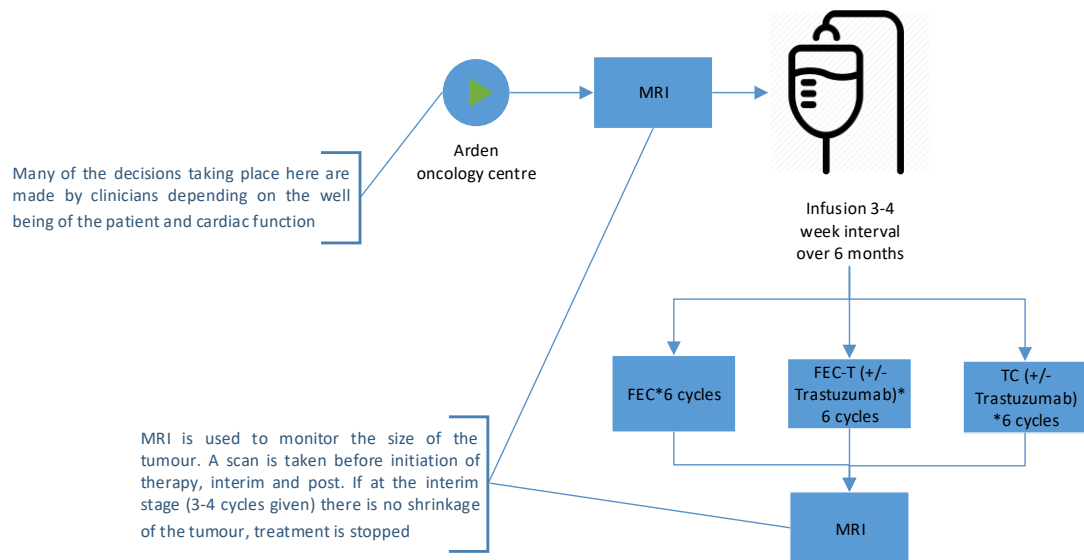
The majority of decisions around the subsequent steps in breast management are made at the diagnostic MDT. The outcomes of the triple assessment are discussed and clinicians make collective recommendations that can be discussed with patients. All decisions and possible outcomes at this stage of the pathway are presented in figure 4-11. The MDT decision algorithm overlaps with the surgical pathways for DCIS and IBC developed from reviewing the guidelines, given in figure 4-3 and figure 4-4 respectively. Once again, there were no contradictions between the evidence base and clinical practice except that clinician interview allowed a slightly more detailed illustration of the pathway. A greater amount of material was available on the handling of benign (B2) and inconclusive (B3/B4) cases. B2 cases were not covered by the guidelines, however this can be expected as they are not cancerous cases. B3/B4 cases are referred for repeat biopsy in order to collect a larger sample of breast tissue, the clinicians did not specify the type of breast biopsy but recommend a larger surgical excision if atypia is present to remove all the affected cells. Once more, this was not discussed in the guidelines. The guidelines focus on confirmed cancerous cases, both DCIS and IBC but do not extensively discuss the lower reported breast pathology categories. While guidelines and clinical practice matched for DCIS, for IBC there was greater detail gained from the clinician interviews. In particular, the clinician interviews gave a clearer understanding of the circumstances under which neoadjuvant therapy was recommended since these decisions were identified as an evidence gap when mapping out the evidence based pathway.

Figure 4-11 Diagnostic MDT decision algorithm



As well as understanding the circumstances under which neoadjuvant therapy is recommended, clinicians were able to give an indication of the commonly prescribed chemotherapy regimens included in breast cancer treatment plans at both the neo-adjuvant and adjuvant phase of treatment. Figures 4-12 and 4-13 highlight the use of FEC chemotherapy unless it is contradicted and so cyclophosphamides are recommended.

Figure 4-12 Neoadjuvant therapy



However, FEC is administered with or without Docetaxel (T) and at both a 75mg and 100mg dose; it was not possible to draw out a decision algorithm at a more granular level, describing when each of these regimens is recommended. The breast oncologist emphasized the challenges in adopting a universal approach but rather each patient must be considered independently taking into account breast cancer features, age, presence of co-morbidities, previous cancer management and the treatment response before deciding the regimen to recommend. Chemotherapy decisions are made by clinical oncologists in the oncology centre and treatment involves close monitoring of patient function and response to treatment, the appropriate regimen is continually assessed and modified to reflect the condition of the patient being treated.

The remaining decision and activities around ER, HER2 and radiotherapy in figure 4-13 and the pathway in figure 4-14 for post-surgery MDT decision making are identical to those recommended in the guidelines.

Figure 4-13 Adjuvant therapy

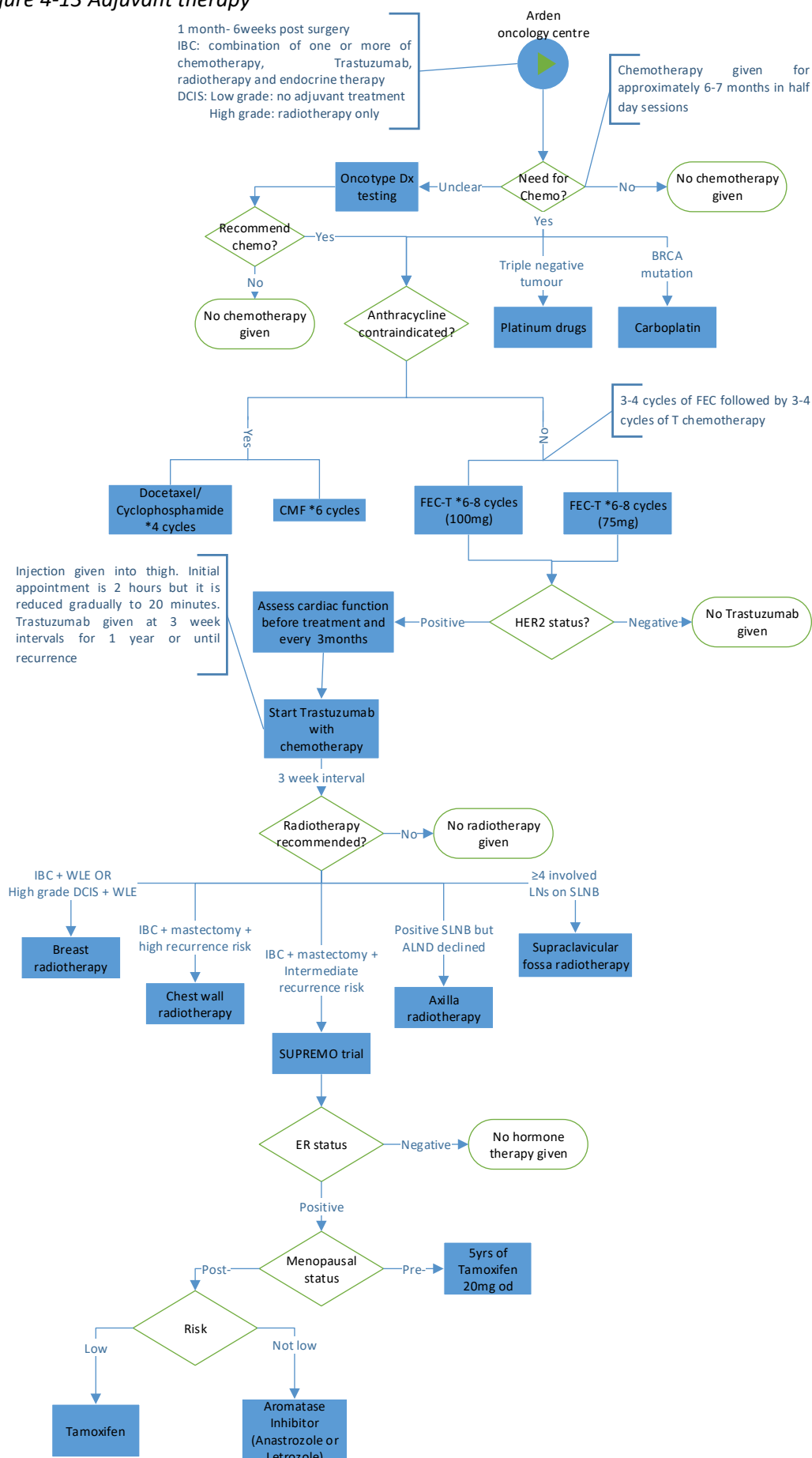
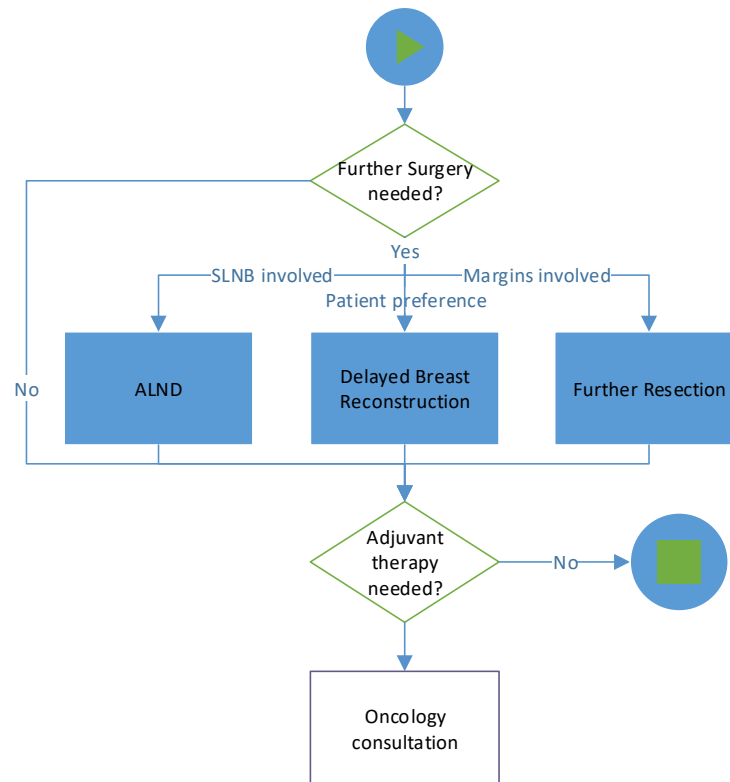


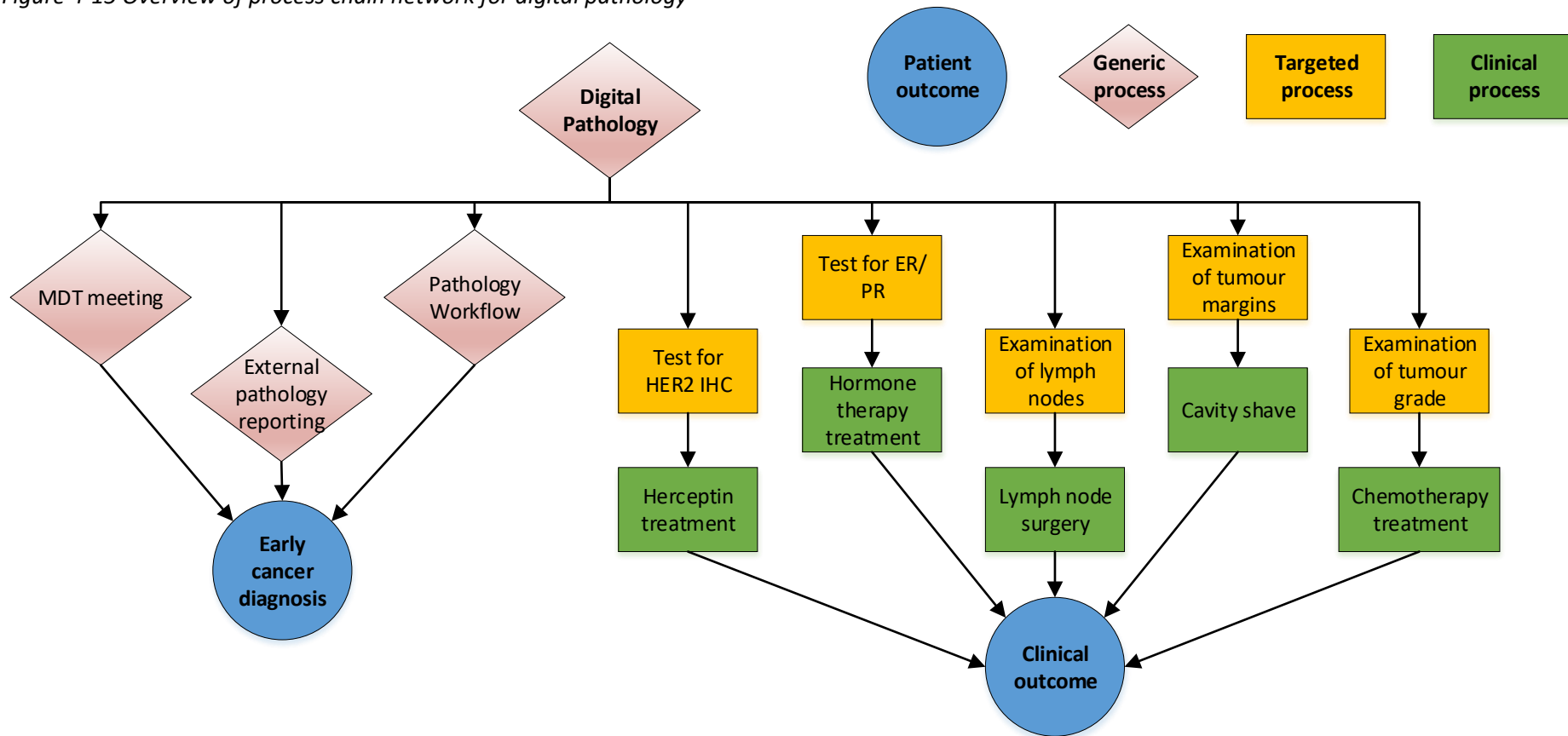
Figure 4-14 Post-surgery MDT



4.4.3 Relationship between digital pathology and breast cancer outcomes

The points of impact of pathology digitisation within the breast cancer pathway learnt by reviewing the literature were presented in chapter 3. These were mapped onto the clinical practice pathway to generate the network of processes from pathology digitisation through to patient outcomes; this is illustrated in figure 4-15. Digital pathology can potentially impact the cancer pathway in two ways, first through gains in accuracy and secondly through gains in speed. The processes for the accuracy gains were straightforward to map onto outcomes for breast cancer patients. Digital pathology can contribute to improving accuracy for a range of targeted processes such as testing or pathological examination of slides which in turn influence clinical processes such as the treatment or surgical intervention recommended which have a direct relation with the patient. To measure the full extent of the impacts of accuracy enhancements within the breast cancer pathway with digitisation, changes in outcomes for HER2 and ER scoring and the examination of lymph nodes, tumour margins and tumour grade would need to be explored. The value of change in accuracy for example by measuring test performance can be mapped onto the appropriate clinical processes as illustrated in figure 4-15 to measure outcomes of these for the patient. The combined benefits of each of the clinical processes can be measured to understand the total benefits for the breast cancer patient due to the digitisation of pathology.

Figure 4-15 Overview of process chain network for digital pathology



Unlike the accuracy gains, the speed gains through digitisation do not feed directly into targeted processes that can be mapped onto a clinical process to measure patient outcomes. Digitisation enhances the efficiency of pathology workflow, external pathology reporting and MDT meetings. These are all within themselves generic processes that each have a diffuse effect across many disease areas.

4.5 Discussion and conclusion

Adherence to evidence based guidelines that translate scientific research into recommendations to support clinicians' decision-making is key to improvements in patient care. Lugtenberg et al. (2009) systematic review evaluated the effectiveness of clinical guidelines; they reported evidence of improvements in both the structure of care and patient outcomes with their implementation. However, in reality guidelines can be over ambitious and difficult to put into practice especially when the appropriate facilities are not in place to support adherence. In addition, clinicians may not always diligently adopt and follow recommendations for numerous reasons including; lack of familiarity, awareness, agreement and motivation (Cabana et al., 1999). As a result, some level of disparity was expected between published breast cancer guidelines and clinical practice but the opposite was found to be true.

The information gained from UHCW clinicians supplemented rather than contradicted the published breast cancer management recommendations. There were no discrepancies between the first version of the pathway based on guidelines alone and the second based on clinical practice. From the standpoint of clinical decision-making, patients at UHCW appear to be receiving good practice evidence based care in the management of their cancer diagnosis and treatment. There is more than one possible justification for this observation:

- UHCW is a large teaching hospital with in-house breast cancer management guidelines that are based on published recommendations. It is evident that they have systems in place for the update, dissemination and implementation of these.
- The nature of the disease area and the continual emphasis on the improvement of cancer outcomes through government targets may influence clinicians' uptake of new research that may not be seen with other disease areas

Even though the findings made on adherence to evidence based practice are positive they may not be generalisable beyond UHCW to smaller hospitals or perhaps even beyond the group of clinicians interviewed, to more junior staff members or clinicians that do not sit on MDT meetings. Furthermore, it is perhaps vital to recognise that the similarities between guidance and practice are based on what clinicians say they would do under certain circumstances rather than what they in fact do which can only be learnt by collecting and analysing individual patient level data, discrepancies between guidance and practice may have arisen if this approach was taken.

The resultant breast pathways highlighted the importance of both guidance and clinician experience when making decisions around the management of disease, this was especially evident in the choices around chemotherapy recommendations. It is difficult to adopt a single approach to disease management for all patients in practice and it is impossible for guidance to cover all types of individual patients with their specific characteristics. Guidelines will provide recommendations for the most typical cases and are not intended to be followed dogmatically without consideration for the patients' wider needs (Green and Piehl, 2003). Thus, it is vital that guidelines are implemented appropriately, together with clinicians' judgements through MDT discussions, especially when more complex and rarer cases are being managed.

Nonetheless, an accurate and comprehensive understanding and depiction of the breast cancer pathways' details from referral to secondary care through to treatment planning based on both evidence and clinical practice is now available for use. Decision analytical models based on part or all of the breast cancer pathway can be developed and used in the evaluation of interventions at any point along the pathway since the basic processes and the relationship between the various activities and the underlying decisions involved have now been summarised in one document. Furthermore, the pathway can be periodically updated to reflect new recommendations but only after confirming feasibility of their application with UHCW clinicians.

The clinical practice pathway can in particular be used to assess the impacts of digital pathology on the overall breast cancer pathway. As described in chapter 3, DP is expected to make multiple improvements in time and accuracy at several points through the breast patients' management pathway. Evaluating each of these independently will not enable the appreciation of all the impacts of digitisation. Through mapping of the information learnt in chapter 3 onto the breast cancer pathway it was possible to summarise the chain of events connecting pathology digitisation to the other activities in the pathway and eventually patient outcomes.

This is the first step in identifying the elements of the pathway that are relevant and should be considered when evaluated the combined impacts of pathology digitisation. While pathology outputs influence some activities within the breast cancer pathway, they do not have a relationship with all of them; these can be eliminated in the evaluation of pathology services allowing simplifications to be made to future modelling exercises. The development

of a decision analytical model that captures and measures the combined impacts of DP is an iterative activity and this is the first step in that process.

In conclusion, the breast cancer pathway has been developed based on evidence and clinical practice for use in the evaluation of interventions along the breast cancer pathway. The pathway allows a clear understanding of events together with interactions between them through decision-making. It is possible to appreciate how modifications in certain activities will influence others and eventually overall patient outcomes.

Chapter 5: Economic evaluation of digital pathology

5.1 Introduction

The breast cancer pathway informed by published recommended guidelines and current clinical practice at UHCW was illustrated and described in chapter 4. This mapping activity provided insight into the relationships between processes within the pathway that cross multiple medical specialities. This useful tool can be used in the initial steps of the evaluation of any intervention along the pathway by first understanding the chain of events that link interventions to patient outcomes.

A combination of general time savings for the pathology lab and accuracy gains for breast pathology examinations have been reported in the published literature and were described in chapter 3. It is important to understand how each of these can contribute to time to diagnosis and treatment selection for cancer patients to achieve government waiting time targets on early diagnosis and improved disease management.

The evolution of cancer waiting time targets set by the Department of Health in England has previously been outlined in chapter 1. Digital pathology activities fall in the period between arrival in secondary care and diagnosis, currently there is no explicit cancer waiting time target that is specific to this time period. The 62 day target from GP referral to the patient's first treatment (Department of Health, 2007) encompasses time to diagnosis but also includes time to treatment. The most recent guidelines that are still being rolled out and have not yet been fully implemented, stipulate replacing the two week waiting time target with a four week time to diagnosis from referral in 95% of cases and two week time to diagnosis for 50% of cancer cases (Independent Cancer Taskforce, 2015). Pathology digitisation can play a role in achieving these waiting time targets for breast cancer patients.

Knowledge of the overall pathway in chapter 4 together with an understanding of the effects of pathology digitisation for the breast cancer pathway in chapter 3 facilitated the illustration of the process networks for digital pathology, figure 4-15. The chain of events describes how the various effects of pathology digitisation feed-into targeted processes which in turn feed-into clinical processes and eventually the desired outcomes. The efficiency gains reported with digitisation can in theory result in patients being diagnosed, treated and discharged earlier while the accuracy gains have the potential to alter the treatment packages prescribed and therefore patient outcomes.

While these impacts are welcome and theoretically support care providers in meeting cancer waiting time targets as well as better planning treatment they have not been measured in

combination to value their collective benefit to breast cancer patients when compared to using standard non-digital pathology methods in the lab.

In addition to the benefits described when adapting the pathology lab to digital systems, costs consequences of pathology digitisation are expected. High upfront investment and regular maintenance costs can make digital pathology appear unattractive to NHS budget-holders. Though savings in resource use costs owing to treatment avoided or the reduction in test requests due to accuracy gains may also be experienced within the lab.

The cost of investment and maintenance can be weighed against changes in the cost of treatment and benefit gains such as time savings in the pathology workflow, improved efficiency in running MDT meetings and better test result accuracy. If the total cost of managing each breast cancer patient in secondary care is reduced with digital pathology when compared to conventional pathology, there is a case to support the adoption of digital pathology at least from the breast cancer care perspective if gains in outcomes are proven, digital pathology will dominate conventional microscopes in the lab. Where the cost per breast cancer patient for digital pathway is greater than that under the conventional pathology pathway, a cost-effectiveness ratio can be generated to determine cost-effectiveness and support decision-makers' investment choices. Costs and outcomes must be measured alongside each other through an economic evaluation in order to determine the suitability of investment in digital pathology.

To measure and quantify all the expected gains in the breast cancer pathway as described in the process chains in chapter 4, several steps were taken to transform the breast cancer pathway illustrated in chapter 4 into a software based decision analytical model to be used for the economic evaluation of digital pathology.

5.1.1 Decision modelling for economic evaluation

Alternative to immediate investment in high-end technologies, decision modelling allows the synthesis of the available primary evidence on effects and costs within an analytical framework (Philips et al., 2006). The model is altered and modified to reflect digital pathology, allowing the comparison and analysis of outcomes from the model prior and post pathology digitisation. The results of the analyses are used by decision makers to make judgements on the adoption of new technologies (Petrou and Gray, 2011) such as digital pathology, in resource limited systems such as the NHS.

There are various models that can be developed and used for healthcare decision making, these include decision trees, markov models, microsimulation models, dynamic models and

discrete event simulation models (Kuntz et al., 2013). The choice of model is largely driven by the clinical question and the presence of interactions between the individuals and systems that are modelled (Barton et al., 2004). Digital pathology's main influences are on efficiency, consequently time to diagnosis and treatment and so an approach that explicitly incorporates time into the modelling process is needed. Discrete event simulation (DES) models will simulate the exact time of each activity in the pathway (O'Mahony et al., 2015). This approach to modelling will enable the development of an understanding of changes in time in the patient's pathway as a result of changes in time seen in the pathology lab due to efficiency improvements with digitisation.

Furthermore, progression of patients through the breast cancer pathway is limited by resource availability, particularly pathology staff working hours and the frequency and length of MDT meetings. DES modelling can capture these constraints, blocking the progression of individuals through the model when resources are limited and capturing the changes that pathology digitisation can bring. An additional challenge to modelling the breast cancer pathway is the level of complexity that needs to be incorporated within the model to fully represent all the activities and decisions that are made within the pathway. Various tests are carried out and results reported as patients move through the breast cancer pathway, DES allows the assigning of attributes to individual patients in the model while concurrently integrating an underlying decision algorithm that controls how patients travel through the model based on their test results (Karnon et al., 2012). DES models are designed to imitate the processes of complex systems, in this case the breast cancer pathway. The pathway is represented in the model and events at different time points can be altered (Jacobson et al., 2006) to reflect the changes predicted with digitisation and analyse their impacts.

Unlike the majority of healthcare interventions, pathology digitisation has more than one outcome of interest, including time spent in secondary care, recommended treatment plans and the number of MDT discussions per case. It is difficult to summarise these into one comprehensive unit of effect to carry out a cost-effectiveness or cost-utility analysis. The research into the impacts of pathology digitisation on both patient care and pathology workflow are still in the early stages as is evident through the quality and size of the literature described in chapter 3. The systematic review in chapter 2 did not identify any economic evaluations of digital pathology or an evaluation of other pathology process interventions that developed a DES model as a vehicle for their economic evaluation. The purpose of an economic analysis at this stage is to explore and highlight the effects of digital pathology on the breast cancer pathway that are of the greatest value.

Cost-consequence analyses (CCA) are a suitable approach for carrying out this type of economic analysis. They allow the presentation of all costs and outcomes of an intervention in a disaggregated format providing decision makers with the most comprehensive information on the value of an intervention (Mauskopf et al., 1998). This exploratory economic analysis will act as a platform from which to identify and justify additional targeted research into both the effects and thus the cost-effectiveness of pathology digitisation.

To measure and value the changes in both accuracy and time to the breast cancer pathway when pathology labs move from conventional approaches to more digital techniques for examination and reporting on tissue specimens, a DES model was developed based on the breast cancer pathway described in chapter 4. The model is founded on both research and clinical practice and has been modified through simplifications to particularly represent pathology activities and the interactions they have with other services along the breast cancer pathway. Digital pathology's influences on the breast cancer patient are indirect and the model allows the linkage between these and the patient. The model was initially populated with parameters that reflect the use of conventional microscopes for pathology reporting. Adjustments were made to the relevant model parameters during experimentation to reflect digitisation of pathology services and compare changes in costs and outcomes for the breast cancer patient with each approach.

5.2 Aims

There are two aims to this chapter:

1. To develop a DES model of the breast cancer pathway that can be manipulated for use in the evaluation of pathology interventions
2. To carry out a cost-consequence analysis of pathology digitisation when compared to conventional microscopes for the examination and reporting on breast tissue specimens. A list of disaggregated costs and benefits for the breast cancer pathway is presented for each pathology approach for comparison

5.3 Methods

This section of the chapter is split into two parts. Firstly, the steps taken to develop a DES model of the breast cancer pathway that can be used in the evaluation of pathology interventions are described. This includes the steps in producing the model structure and data collection methods adopted to inform the models' time, cost and accuracy parameters so that the model reflects the standard practice of conventional microscope use in the pathology lab.

The second part of this section describes the steps taken for the economic evaluation of digital pathology. A description of how the experimental factors within the DES model structure are manipulated to reflect pathology lab digitisation are given. Followed by a summary of the cost and benefit outputs that are compared across both CM and DP systems.

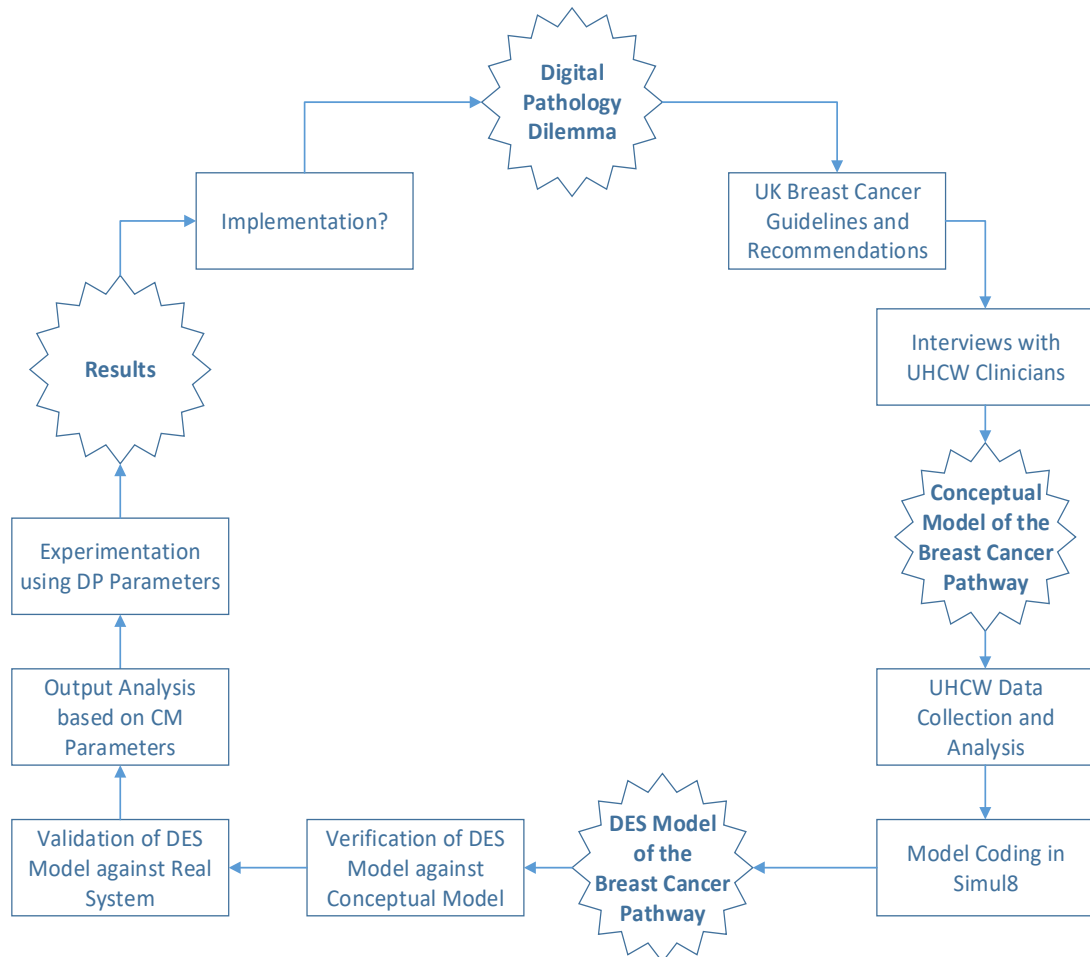
5.3.1 DES model development

It is vital models are developed to accurately reflect the real system they are based upon with all underlying assumptions and simplifications justified satisfactorily. Any policy changes built on erroneous outcomes as a result of inadequate modelling will not only have a negative cost impact but in the healthcare setting maybe detrimental to a patients' wellbeing (Jacobson et al., 2006). Due to the complex nature of DES modelling a step-by-step approach as detailed in figure 5-1 was adopted to develop the model that is used in the analysis of digital pathology. **Error! Reference source not found.** 5-1 is adapted from information learnt from a combination of resources including Law (2003), Carson et al. (2005), Hoad (2016) and Robinson (2004).

The steps outlined in figure 5-1 were followed. The review of UK guidelines and interviews with UHCW clinicians have already taken place and have been described in chapter 4. The resultant breast cancer clinical practice pathway was used to inform the conceptual model that is transformed into the DES model to be used in the evaluation of pathology interventions. The clinical practice pathway represents the full spectrum of events along the breast cancer pathway, simplifications and assumptions were made to this to design the conceptual model structure of the DES model so that it was relevant to the research question and all unnecessary processes are removed. This was followed by the collection of the appropriate time, cost and probability data from UHCW and other sources to inform the model parameters so that the model reflects the standard use of conventional microscopes in the lab. The conceptual model and the data collected were used to develop the DES model using Simul8 software. Verification and validation processes were carried out to check the

model was developed correctly and confirm that it can be used in the evaluation of pathology interventions.

Figure 5-1 Steps in DES model development



5.3.1.1 The conceptual model

“The conceptual model is a non-software specific description of the computer simulation model (that will be, is or has been developed), describing the objectives, inputs, outputs, content, assumptions and simplifications of the model.” (Robinson, 2008, p.282)

Conceptual model development is the first phase of DES modelling; this is a schematic of the breast cancer pathway that is relevant to the model objectives. The purpose of the final DES model is to explore the implications of digital pathology for the breast cancer patients’ management pathway. To ensure the digital pathology experimental factors as well as the model outputs are modelled appropriately they were identified at this stage of model development and explicitly incorporated into the conceptual model so that the model objectives and the necessary analyses to be undertaken were at the centre of all model development.

An overview of the process chain network for digital pathology was given in figure 4-15 in chapter 4, this outlines the processes within the breast cancer pathway that are expected to be influenced by digitisation and must be represented within the DES model for evaluation. However, primary data comparing CM and DP outputs for each of these processes is not yet available as reported in chapter 3. Whilst the conceptual model made sure that all processes were included they will not all be explored in the experimentation phase of the evaluation. Time savings in the pathology pathway, increased productivity during MDT meetings, improvements in HER2 accuracy and changes to the tumour grade reported are the main outcomes of digitisation that were explored, these were the experimental factors of interest. The remaining processes in figure 4-15 were still represented in the model so that as primary data becomes available for each of these experimental factors the evaluation can be updated to explore further the impact of digitisation.

The experimental factors listed will have a knock-on effect on time in the pathway, the treatment plans prescribed, the number of MDT discussions per patient as well as NHS costs. These are the model outputs of interest.

In essence, the clinical practice pathway described in chapter 4 is the foundation of the conceptual model. Whilst it embodies all stages of the breast cancer pathway, relationships do not exist between the experimental factors and the full scope of activities represented. The pathway was simplified to incorporate sufficient detail to capture the activities that are sensitive to the experimental factors, avoiding unnecessary complex modelling. The model structure represents the network of pathology interactions within the breast cancer pathway in its simplest form, ensuring that model accuracy is not undermined whilst meeting the research objectives.

Conceptual model development is an iterative process, so whilst the initial version was based on information learnt in chapter 4 further information about the pathway was derived from the data collection activities carried out at UHCW. The majority of the data collected was used to inform the model parameters and is discussed in the next sections of this chapter however in some instances the data highlighted features of the clinical practice pathway that were not modelled or had not been previously considered. This was particularly around the types and features of the tissue specimens removed rather than the pathway structure itself. For ease of reporting, where information not previously known was learnt about the breast cancer pathway through data collection it is described and justification for its omission or inclusion are given in this section.

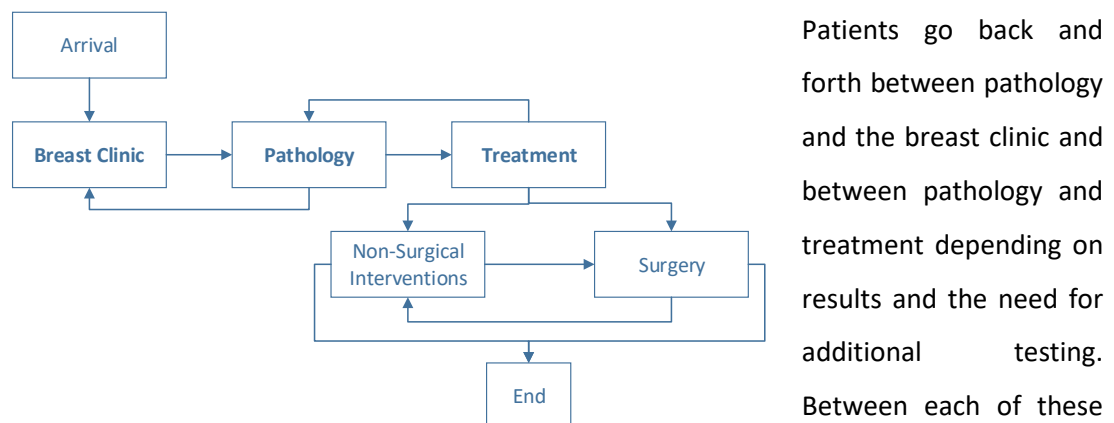
The full breast cancer pathway in chapter 4 was reviewed and any sections considered not relevant to the ultimate model purpose were removed. These are clearly listed and justifications for each omission or assumption are given. The conceptual model is a simpler version of the clinical practice pathway designed to meet the evaluation objectives.

5.3.1.1.1 Simplifications and Assumptions

The pathway of a patient with suspected breast cancer referred to secondary care for investigation can be summarised into three phases as illustrated in figure 5-2:

- (1) Breast Clinic: Patients arrive here for triple assessment including the removal of biopsy samples
- (2) Pathology: The initial biopsy and other specimens removed are taken to the pathology labs for investigation.
- (3) Treatment: The pathology results guide treatment planning, including both surgical and non-surgical interventions where recommended.

Figure 5-2 Summary breast cancer pathway through secondary care



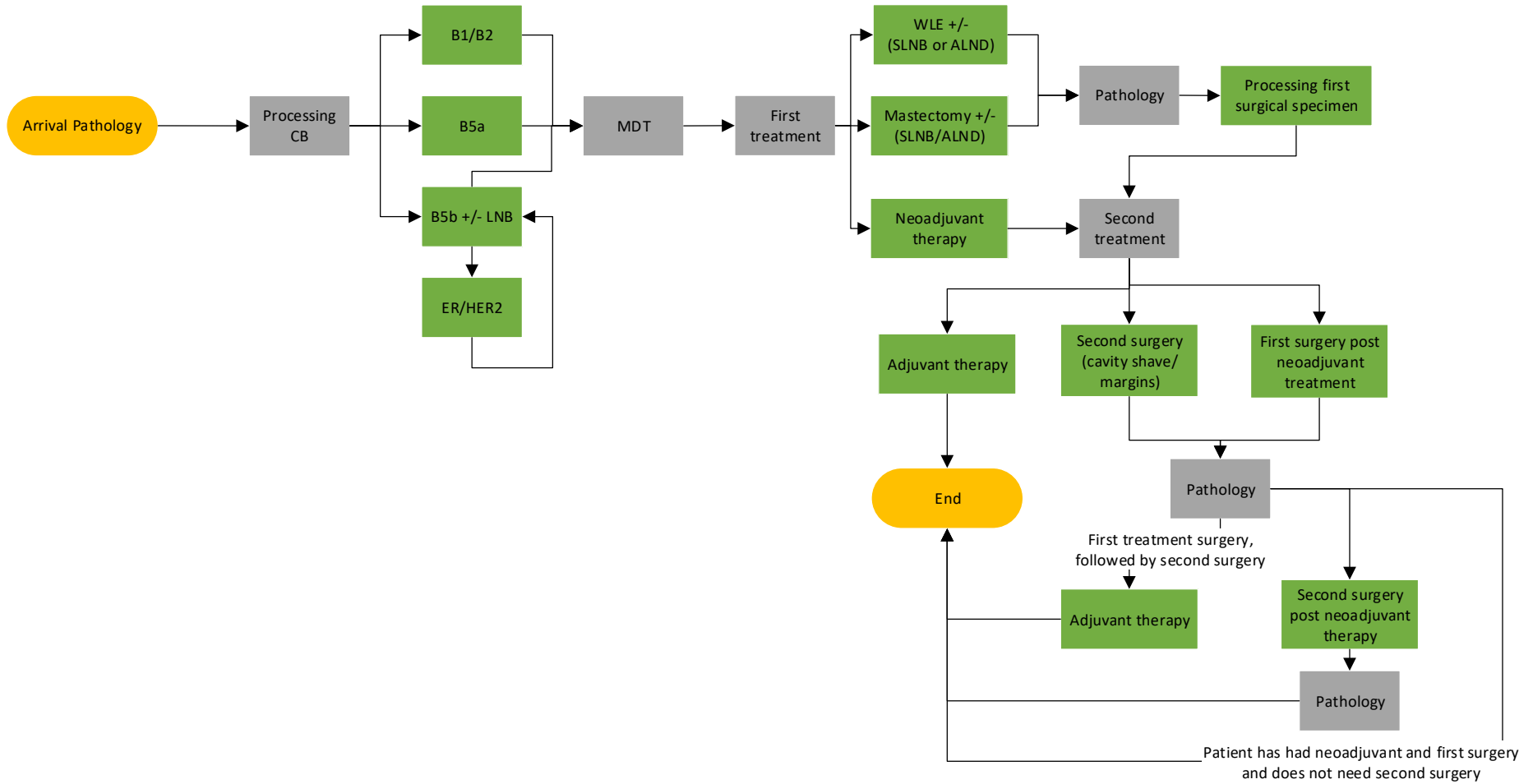
MDT meetings and OPAs are taking place to decide the course of action and communicate this to the patient. This is just an overview, in reality each one of these phases is made up of a series of activities and decision points. The challenge is to identify which elements of the initial schematic of the breast cancer pathway are to be included in the conceptual model for conversion into the DES model. A summary of all model simplifications and omissions from the breast cancer pathway are listed in table 5-1 and an overview of basic structure of the conceptual model is in figure 5-3. This is a representation of the elements of the original breast cancer pathway that are eventually included in the model, in reality the conceptual model is a lot more detailed than this.

Table 5-1 Summary of simplifications in DES model

Simplifications/omissions	Justification
Breast clinic	
The breast clinic and triple assessment activities are not represented	They do not interact with and therefore are not sensitive to changes in pathology
Patients are managed in the same way regardless of source of referral	Differences in patient management due to source of referral only arise in the breast clinic and this not modelled.
Breast clinic and the activities within it are not modelled	
Pathology – First specimen	
Breast reductions are not represented	They are not part of the cancer pathway, cancer for these cases has previously been managed and patient has come back for procedure to manage risk
Soft tissue specimens are not represented	Where cancer is suspected, a core biopsy is always carried out
Assume only one breast investigated for all patients	Patients are managed according to breast with most serious result
Assume only one core biopsy taken for all patients	Patients are managed according to biopsy with most serious result
First specimen is a core biopsy of one lesion of one breast	
Pathology – CB results	
B1 and B2 core biopsies cases are merged into one group	These cases are managed in the same way
B3 and B4 cases are not represented	These cases always go on to have a second biopsy that eventually gives a B1, B2, B5a or B5b result. Their incidence is not influenced by pathology
Core biopsies have one of 3 diagnoses B1/B2, B5a or B5b	

Treatment	
Assume that all patients follow through their treatment plans	Beyond the scope of the study to model interim changes in treatment plans
Assume all patients are early breast cancer cases	Beyond the scope of the study to model metastatic and recurrent cases as their management plans differ to early breast cancer cases
Delayed breast reconstruction omitted	This activity is not influenced by pathology
Radiotherapy activities	This activity is not influenced by pathology
Excision of B1/B2 lumps not modelled	Not restricted by cancer waiting time targets
The use of MRI to monitor the size of tumour during Herceptin treatment	The model previously assumes that patients follow through their treatment plans MRI guides if treatment should be continued so this activity is no longer relevant
Chemotherapy regimens summarised in one activity	Pathology influences whether chemotherapy is prescribed not the type of chemotherapy regimen
Hormone therapy regimens summarised into one activity	Pathology influences whether hormone therapy is prescribed not the regimen type

Figure 5-3 Overview of the conceptual model



Breast Clinic

1. Triple assessment

Patients are first exposed to pathology services after they have completed their triple assessment in the breast clinic. The biopsy specimen of the unexplained tissue mass is sent to pathology labs for further investigation. All the steps in the breast cancer pathway prior to this are not sensitive to changes in pathology. DP will not influence clinical examination, the imaging options or even methods of biopsy. The impacts of pathology digitisation will start to be seen from the point of arrival of the biopsy specimen at pathology through to treatment design and initiation. With this in mind, it is clear the level of detail found in the clinical practice pathways prior to pathology do not need to be represented in the conceptual model. The model is simplified by removing the arrow from the breast clinic to pathology, omitting all activities involved in triple assessment.

2. Source of referral

Patients arriving in secondary care with suspected breast cancer either have presented to their GP with symptoms or have been recalled from routine assessment as part of the breast-screening programme. Even though there are small differences in the activities and decisions that make up triple assessment for these two groups, once biopsy specimens have been removed, all patients are managed in the same way regardless of their source of referral. As the activities representing triple assessment have already been removed from the model, there is no need to differentiate breast cases by their source of referral. Breast cancer cases arriving in the conceptual model are simplified and treated as one homogenous group. The model begins with the arrival of the first biopsy specimen for each breast case at the pathology lab.

Pathology

1. First specimen

The evidence and clinical practice pathways point to the core biopsy as the first tissue specimen to be removed from the patient on arrival in secondary care for investigation. In the UHCW dataset, over 85% of patients had a core biopsy (CB) of the breast removed with or without a lymph node biopsy (LNB) as their initial tissue specimen, this matches the breast cancer pathways developed in chapter 4. However, the remaining 14% of patients have other initial tissue samples, including skin biopsies from the areola and/or nipple, soft tissue specimens taken from the chest wall and breast reduction tissue.

a. Breast reductions

Reductions are indicative of a previous episode of breast cancer; here the patient has returned to either have their non-cancerous breast operated on to bring its size in line with the cancerous breast that has undergone a previous surgical procedure. For some patients a breast reduction is carried out to reduce the risk of recurrence even though the breast operated on does not yet show any signs of malignancy. This group of patients are excluded; in most cases there is a long period between first diagnosis with breast cancer and the breast reduction procedure and so for many of these patients information is not available for their previous cancer episode. It is not necessary for the model to represent these cases since breast reductions are in reality not the first tissue specimen removed in their pathway.

b. Soft tissue specimens

The soft tissue specimens in the dataset are extracted as a result of a suspected lipoma. These are benign lumps which are typically completely excised. Since they are not suspected cancers and do not follow through the breast cancer pathway there is no need for them to be represented in the model. For the skin biopsies in the dataset, only one case was confirmed to be cancer and this was only after a core biopsy was performed when signs of malignancy were seen on examination of the skin biopsy in pathology. The core biopsy for this patient can be considered their initial specimen in the pathway and skin biopsies are not represented in the model.

c. Two breasts investigated

The UHCW dataset showed the vast majority of patients undergo investigation in one breast alone, however approximately 3% of the patients in the dataset had biopsies taken from each breast due to palpable lumps in both. For ease of modelling and simplification purposes these patients were assumed to only have one breast investigated but will be processed through the model according to the breast with the more serious diagnosis as this is what occurs in practice. The inconsistencies between the model and reality are associated with the cost and time needed to process two core biopsies rather than one and the cost and time of surgery if recommended for both breasts rather than just one; the remaining aspects of these patients pathway will be identical to patients that have only one breast under investigation.

d. Two core biopsies

Similarly, some patients have more than one core biopsy lesion removed from the same breast where there are several palpable lumps. In most cases these have the same result but

where they differed, the patient is treated according to the more serious outcome. Once more for simplification purposes, results for more than one lesion in the same breast have been combined to give an overall breast result that is included for the particular patient. The model assumes a diagnosis is made with one core biopsy which is labelled and processed through the model according to the lesion with the more serious diagnosis. The irregularities between the model and reality are associated with the cost and time to process more than one core biopsy in pathology rather than one lesion as is the case in the model.

In summary, it is maintained that the breast cancer case arriving at the start point of the model has had a single core biopsy removed with or without a lymph node biopsy from one breast. All other specimen types are not considered.

2. CB result

Pathologists classify breast core biopsies with a B1 to B5 result as outlined in table 5-2 below (Andreu et al., 2007).

Table 5-2 Summary of pathology CB results

B1	Normal
B2	Benign
B3	Lesion of uncertain malignant potential
B4	Suspicious
B5a	Malignant – Insitu
B5b	Malignant – Invasive

The majority of initial biopsies in the dataset were determined to B1, B2, B5a or B5b. In a minority of cases a diagnosis is more difficult and the lesion is determined B3 or B4.

a. B1/B2 cases

B1/B2 core biopsies represent either normal or benign tissue, the benign classification embodies a diverse group of lesions ranging from cysts to fibroadenomas (Guray and Sahin, 2006). In most cases these lesions do not require intervention by clinicians but there are some instances where surgical removal of the lump is recommended. In the UHCW dataset 90% of patients' with B1/B2 core biopsies are discharged post MDT and for 10% of cases MDT advise the excision of the benign lump.

b. B5a cases

A B5a result on CB denotes an insitu lesion, ductal carcinoma insitu (DCIS). These lesions are found in the milk ducts and have not spread to other parts of the breast (Ernster and Barclay, 1997). DCIS can appear alongside an invasive lesion and in such a case it is treated as invasive breast cancer (IBC). These lesions are straightforward to diagnose and manage; at this stage it is known that they are malignant but have not spread into the breast. B5a lesions are generally managed by surgery to remove the tumour; either a mastectomy or wide local excision (WLE) depending on the size of the area of the breast affected (Health and Excellence, 2009).

c. B5b cases

A B5b result on a CB indicates an invasive lesion. Unlike DCIS, IBC has spread from the ducts and into the surrounding breast tissue. These cancers are also more likely to be involved with lymph nodes indicating potential metastasis through the lymphatic system (Rahman and Mohammed, 2015). Resection of IBC tumours is at the core of the management strategy for this category. Like DCIS tumours; patients will either undergo a mastectomy or a WLE, though some patients that meet the criteria for WLE will opt for a mastectomy instead (Rostas and Dyess, 2011). As well as surgical resection these patients will also have an assessment of their lymph nodes by way of a sentinel lymph node biopsy (SLNB) to determine if the cancer has spread and thus guide subsequent treatment (Rostas and Dyess, 2011).

B1, B2 and B5 lesions are relatively straightforward to manage (Pinder and Reis-Filho, 2007). B1 and B2 lesions are mostly discharged and B5 will be discussed by MDT for treatment management. The delay arises around B3 and B4 lesions as these will require repeat assessment to make a definitive diagnosis and plan treatment accordingly (Pinder and Reis-Filho, 2007).

d. B3/B4 cases

The B3 category includes a diverse range of lesions of differing levels of possible malignant potential (Strachan et al., 2016); these include:

- Epithelial Atypia
- Atypical Intraductal Proliferation
- In Situ Lobular Neoplasia
- Papillomas
- Radial Scars

- Fibroepithelial Lesions

(Strachan et al., 2016)

B4 lesions are highly suspicious of malignancy; either DCIS or IBC, but a definite diagnosis cannot be made for various reasons (Bilous, 2010b). Both B3 and B4 lesions are dealt with by further investigation, explaining the arrow in figure 5-2 from pathology back to the breast clinic. This is normally a diagnostic excision biopsy (DEB) or a vacora biopsy (VB). A vacora removes more tissue per sample (Ames and Britton, 2011) compared to a core biopsy whilst a diagnostic excision involves the removal of the complete palpable lesion.

In the UHCW dataset, 15 patients had an initial B3/B4 result and required further investigation to make a final diagnosis and initiate their treatment. They underwent either a vacora or a diagnostic excision, the results of these biopsies for each patient are given in table 5-3. Only one patient (ID 80) is found to be cancerous when the second specimen is analysed. The remaining all have benign results and two patients; 108 and 234 had no further investigations.

The need for repeat biopsy is largely due to unusual breast cancer types or when the core biopsy specimen appears to not be completely representative of a larger lesion seen on imaging and so further sampling is needed (Bilous, 2010a). DP will not play a role in overcoming the challenge of inadequate tissue specimens as this is related to biopsy techniques rather than specimen analysis.

Table 5-3 Management of B3/B4 core biopsies in the dataset

ID	CB	Comments	VB	DEB
9	B3	Spindle Cell		Spindle Cell
17	B3	2 Lesions (B1 & B3)	B2	
19	B3	Sclerosed Papillary Lesion without atypia		Benign
44	B3	Intraductal Papilloma		Benign
68	B3	Papillary Lesion no atypia		Normal
104	B3	Infarcted papillary lesion + atypia amounting to FEA/ADH		Low Grade DCIS
106	B3	Part of a papillary lesion without atypia + a focus of atypical intraductal proliferation	B1/B2	Benign
108	B3	Intraductal Papilloma		
146	B3	Focal Epithelial Atypia		Benign
157	B3	Atypical Intraductal Proliferation	B3	Benign

162	B3	Atypical Intraductal Proliferation	B2	
166	B3	Spindle cell lesion, likely myofibroblastoma		Fat only
211	B3	Atypical Intraductal Proliferation	B2	
234	B3			
80	B4		B5a	

The proportion of CBs classified as B3/B4 will not be influenced by pathology. The model is simplified so that all breast cases at the start of the model are assumed to have their definitive CB result assigned on their initial specimen. The route back to the breast unit from pathology in figure 5-2 is omitted; the breast clinic is now not represented in the conceptual model. Furthermore, core biopsies with a B1 or B2 result are managed in a similar manner; these are merged into one group in the model. In summary CBs at the start point of the model have one of three diagnoses; B1/B2, B5a or B5b.

Treatment

The UHCW dataset shows that treatment plans for some patients will not match clinical practice as described by the clinician's interviewed, either from outset or if adjustments to plans are made once treatment is initiated. This can be due to a number of reasons including; the experience of side effects, presence of co-morbidities, non-response or refusal of treatment by the patient. These external factors are not influenced by pathology and it is beyond the scope of the evaluation to capture these in the model. For simplification purposes the model assumes all patients follow through their care plans as would be recommended in the breast cancer clinical practice pathway without taking into consideration any external factors beyond disease characteristics that may influence the treatment decisions made.

The additional factor to take note of at this stage of the modelling process is breast cancer recurrence. Representing these cases in the model adds further layers of complexity; these patients have already received a combination of both surgical and systemic therapy with their first round of treatment. Their treatment management differs to cases presenting with breast cancer for the first time and they have not been covered by the guidelines used to develop the pathway in chapter 4. The model was developed with early breast cancer cases in mind and does not represent the pathways of advanced or recurrent breast cancer cases,

for this reasons all suspected breast cancer biopsies at the start point of the model are related to early breast cancers only.

1. Surgery

All surgical activities in the clinical practice pathway are kept in the conceptual model except for delayed breast reconstruction. This procedure is largely down to patient preference and is not influenced by pathology outcomes. The model is simplified by removing it.

2. Radiotherapy

Recommendations for radiotherapy are guided by the surgical procedure and the LN status reported in the pathology report. It was previously highlighted in chapter 4 that LN evaluation and reporting is fairly accurate with conventional microscope. As long as digitisation does not undermine accuracy there will be no impact on the prescription of radiotherapy regimens. The model was simplified by removing all radiotherapy activities.

3. Systematic therapy

Several simplifications are made to systemic therapy. Firstly, both the use of MRI to monitor tumour size during the course of neoadjuvant therapy and the assessment of cardiac function during Herceptin treatment are not included in the model. Based on the outcomes of both activities, clinicians modify treatment accordingly, but as the model will assume that patients follow through and complete their initial treatment plans it is not necessary to model either of these activities. Furthermore, neither interact with pathology services.

Secondly, as described in previous chapters there are several options for chemotherapy drug regimens and dosage in breast cancer. For simplification purposes due to the excessive complexity of attempting to model each of these, chemotherapy will be represented by one activity in the model. The same also applies for hormone therapy. Pathology results influence the prescription or not of these treatments whereas other patient factors, external to pathology will influence the specific drug to be prescribed.

4. MDT

Observations of hospital practice and knowledge of pathology processes gained through data collection at UHCW indicated that pathology reports are in some cases added to MDT lists for discussion prior to all test results becoming available. This scenario was not seen for CBs with either B5a or B1/B2 status since only the biopsy diagnosis is reported with no further testing needed. In most cases theses were incomplete reports for B5b CBs where other test results are not yet available especially if HER2 FISH testing is being outsourced due to an

equivocal IHC outcome. Rather than wait for a complete pathology report which includes the results of all supplementary tests, the cases are sent to MDT for discussion then returned to the pathology lab for the missing results to be included before being discussed once again at MDT. Alongside MDT meetings, OPAs are pre-scheduled to communicate the agreed care plan with the patient on the day of or day following the MDT meeting where their case is due to be reviewed.

In essence, a case may appear at more than one MDT meeting before a final treatment plan can be proposed resulting in both a waste in clinician time and an inconvenience to patients who are going back and forth for hospital appointments without being given a comprehensive explanation of their diagnosis. Ideally all test results should be available at the first MDT discussion so that results and treatment plans can be discussed with patients at the OPA.

DP is set to improve pathology workflow and consequently reduce the time to test results becoming available. In theory, this will reduce the number of breast cases arriving at MDT meetings with incomplete reports. The impact of DP on the number of times a breast case passes through MDT is one of the investigations explored in the model manipulation stage later on in this chapter.

5.3.1.2 DES model

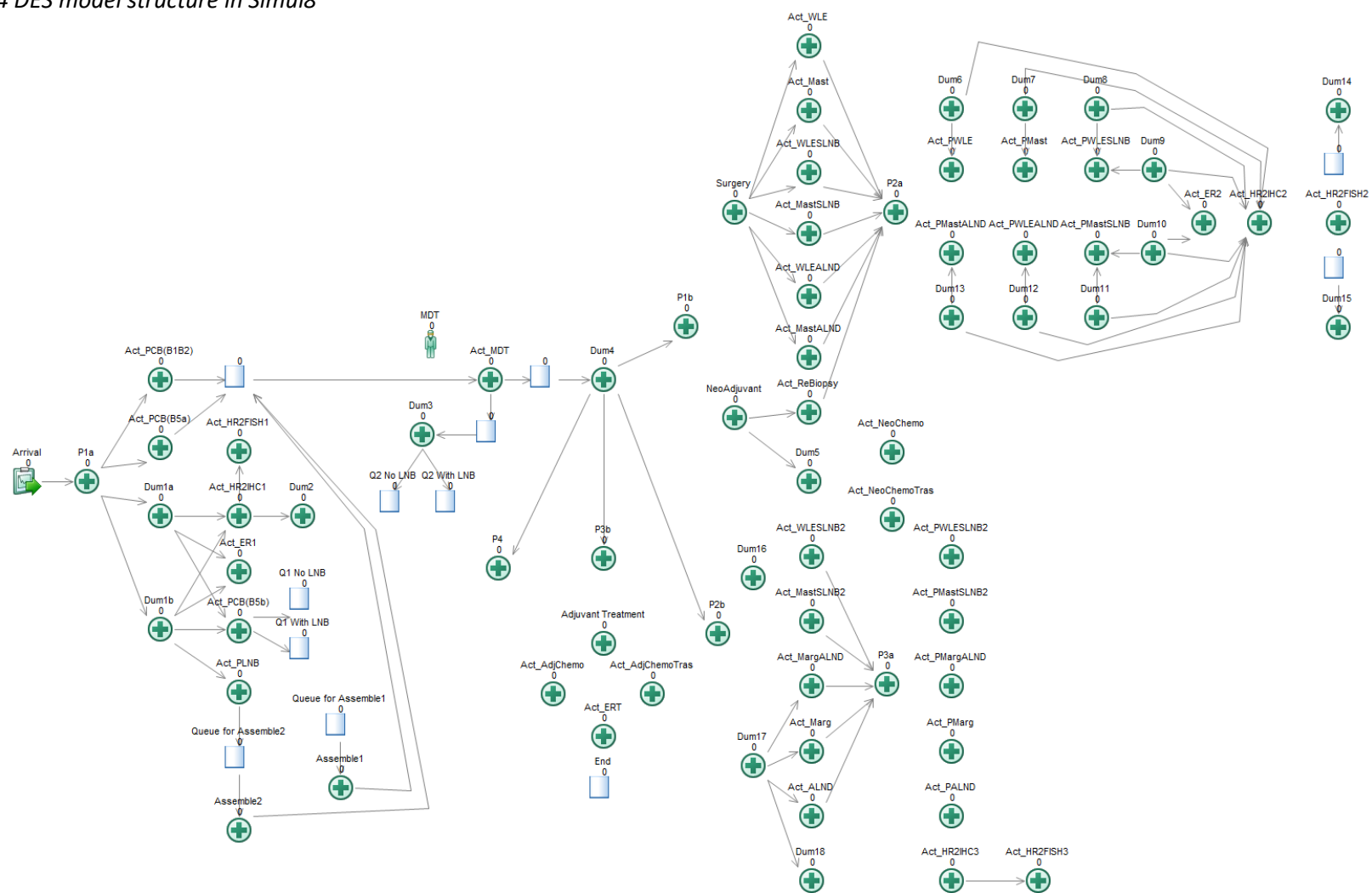
To capture all the time and accuracy consequences of pathology digitisation a DES model was developed using Simul8 Professional software (Simul8, 2019), this is based on the conceptual model described so far, and an overview of this is given in figure 5-3.

A step-by-step approach was taken to develop the model in Simul8:

1. Model assembly

The initial structure was built incorporating all pathology outputs and routing decisions that are made based on these. This was completed using combinations of batching, labels, queues, routing and visual logic code in Simul8 to move entities (breast cases) appropriately through the activities in the model. A description of the Simul8 building blocks is given in appendix C and an illustration of the Simul8 model structure is given in figure 5-4.

Figure 5-4 DES model structure in Simul8



2. Verification

Verification makes sure that the visual logic codes input are correct so that work items are routed through the pathway as described in the conceptual model. Using the step function in Simul8 software allows the tracing of any chosen entity using its unique ID label, this is continually done with the addition of any new code or routing logic during the development phase. Following the addition of any new visual logic code, label updates or routing function at an activity, the model is run and the queue contents are examined to ensure work items are being routed correctly and labels are being updated appropriately. This is an ongoing process and has been carried out thoroughly with every step of model development ensuring the model is running and updating as expected. Owing to continual verification; on completion of the model structure, breast cases are routed in an identical way to that described in the conceptual model with all bugs and errors removed.

3. Model parameters

Once the model structure was completed and all entities were passing through the model, appropriately reflecting the conceptual model, time distributions and costs were added to the DES model at the appropriate activities. A summary of all model parameters and sources is given in the next part of this section.

4. Validation

As well as verification it is essential the model is validated by comparing it to the real system (Robinson, 1997); in this case the breast cancer pathway at UHCW. If the inputs to the model are the same as reality then the outputs from the model should also be reasonably similar to real data. The final step involved validation of the DES model against the real system (UHCW) by comparing model outputs against UHCW data.

The verification and validation processes of developing a DES model are essential in determining whether the model developed is similar enough to the real system for it to be used in meeting the study objectives (Sargent, 2009). Verification focuses on determining the accuracy of the DES model in relation to the conceptual model and validation centres on how close the model represents the true system (Cook and Skinner, 2005).

5.3.1.3 Data collection

From the conceptual model, it was possible to identify the data required in order to populate the DES model. The model parameters can be split into three categories:

1. Probability data

2. Time data
3. Cost data

The first two categories were collected from and are based on UHCW data. The cost data was largely based on the published literature with some data points taken from UHCW systems. The model parameters reflect clinical practice at UHCW i.e. the use of conventional microscope in the pathology lab.

UHCW data

A list of hospital numbers for breast patients was provided by the pathology department at UHCW. These patients had at least one tissue specimen included in the DP validation study carried out at UHCW by Snead et al. (2015). The study included over 3000 specimens of which 265 were breast specimens removed from 235 patients. These were selected at random from February 2013 to April 2014 by taking record cards for breast pathology specimens from the filing tray, once the pathologist had reported their findings. A breakdown of the number of specimens per patient included in the validation study is presented in table 5-4. These breast tissue specimens were the initial source of all the required data to be collected to inform the model parameters.

Table 5-4 Number of breast tissue specimens per patient

Specimens	Patients	Total
1	210	210
2	20	40
3	5	15
Total		265

The tissue specimens ranged from simple biopsies to more complex surgical procedures. Not all of a patient's tissue samples associated with the breast episode were included in the validation study. For example, a biopsy for patient A is included but they later have a WLE that is not picked up in the random selection of records and vice versa. 210 patients are represented by only one tissue specimen in the study; 20 patients have two and for 5 patients there are three breast tissue specimens included in the validation study. The 235 patients were each given a unique ID number from 1 to 235. Three patients were eventually excluded from the dataset; one was discovered to be a gastric biopsy, most likely included in error. Two patient's records could not be found on CRRS. Data for the remaining 232 patients were

either requested from hospital staff or manually extracted from systems where it was possible to gain access.

1. Requested data

All MDT data were requested for patients with breast tissue specimens included in the validation study. These were extracted by staff at UHCW and included information on the date of meetings, specimen discussed and MDT outcome.

It was not possible to calculate the inter-arrival rate of breast specimens received in the pathology lab from the validation study cases. They did not include all the breast cases arriving at pathology during the study period. The basic information for breast cases arriving at pathology between January 2013 and January 2015 was requested. This included information on specimen type, date of arrival and the diagnosis made. From this information an inter-arrival rate was worked out for use in the model.

2. Manually extracted data

As well as data requested, other relevant information necessary to populate the model was manually extracted. There were two sources for this:

The UHCW central results reporting system (CRRS) was the main source of data, from here information on pathology, surgery and oncology treatment plans were extracted. Data was collected on all pathology specimens related to the patients' breast episode not just the specimen included in the validation study.

The oncology information system at UHCW; MOSAIQ was also accessed through the pharmacy department. Information on patient's drug history and treatment plans was taken from here.

Data were extracted from reports associated with events taking place along the patient's breast cancer pathway that are included in the conceptual model. Table 5-5 presents a summary of all records obtained and their source.

Table 5-5 UHCW data collected

Data Group	Source	Parameters
First appointment	Dendrite System Manager	<ul style="list-style-type: none">• Date of first visit
MDT meetings	Dendrite System Manager	<ul style="list-style-type: none">• Date of meeting• Breast tissue specimen discussed• MDT outcome

Breast pathology reports	Manual CRRS extraction	<p>This includes pathology reports for all breast tissue specimens removed during the patients' course of management in secondary care. Encompassing both the cases included in the validations study as well as other breast specimens. The following information was extracted:</p> <ul style="list-style-type: none"> • Date of specimen receipt • Date of pathology report • Main results reported • Supplementary results reported
Surgery	Manual CRRS extraction	<ul style="list-style-type: none"> • Surgical procedure • Date of surgery
Pharmacy	Clinical letters stored under CRRS AND information from the MOSAIQ database	<ul style="list-style-type: none"> • Chemotherapy regimen details: • Endocrine therapy details: • Herceptin therapy details:

Model probabilities

The majority of model probabilities were estimated from the UHCW data described in table 5-5. A summary of all model probabilities and their sources are given in table 5-6.

Table 5-6 Model probabilities

Model parameter	Value	Probability	Sources
CB result	B1/B2	59%	UHCW data
	B5a	6%	
	B5b without a LNB	26%	
	B5b with a LNB	9%	
Upgrade of B5a case to invasive status on surgery	Upgraded to invasive	26%	Brennan et al. (2011)
	Remain in situ	74%	
Surgery if invasive	WLE	57%	National Cancer Intelligence Network (2006)
	Mastectomy	43%	
Surgery if in situ	WLE	65%	
	Mastectomy	35%	
HER2 IHC results	+3 positive	14%	UHCW data

	+2 equivocal	23%	
	+1 Negative	63%	
HER2 FISH results round 1	Amplified	21%	UHCW data
	Repeat test	72%	
	Non-amplified	7%	
HER2 FISH results round 2	Amplified	50%	UHCW data
	Not-amplified	50%	
ER results	Positive	83%	UHCW data
	Negative	17%	
LNB status	Involved	57%	UHCW data
	Not involved	43%	
Systemic therapy plan	Neoadjuvant treatment	3%	UHCW data
	Adjuvant treatment	97%	
Margin status	Involved	12%	UHCW data
	Not involved	88%	
SLNB status	Involved	13%	UHCW data
	Not involved	87%	
Neo adjuvant surgery	WLE	64%	Mieog et al. (2007)
	Mastectomy	36%	
Grade if HER2 negative	1	30%	UHCW data
	2	44%	
	3	26%	
Grade if HER2 positive	1	0%	UHCW data
	2	33%	
	3	67%	

Time data

One of the main objectives of the simulation model is to capture the time changes in the breast cancer patients' pathway that will be seen with the digitisation of pathology services. Experimentation will investigate how digitisation influences the time to complete tasks in the pathology lab, for this reason all pathology related activities are assigned a time distribution as these will need to be manipulated to meet the main objective of the model. The time estimates that were used in the model were based on UHCW data and reflect the pathway when conventional microscopes are used. For each breast patient included in the

UHCW dataset described earlier, the dates of all activities in their treatment and management pathway were extracted. These data points were used to generate time estimates for the activities that take place along the patients' breast pathway and are represented in the conceptual model. Where possible simplifications were made by combining data points and generating a single time estimate that was later used at more than one activity in the model. If it was considered appropriate to merge data points, for example, the time taken to report on a CB with malignant status versus non-malignant status, two-sample t-tests were used to compare the data. Where there was no difference at the 95% level of significance, the data was merged and one time estimate generated. Where there was evidence of a difference between the activity times, the data points were kept separate and two distributions used at their matching activities. All time estimates as used in the model are reported in table 5-7.

Table 5-7 Model time parameters in days

	Mean (95% CI)
Time to process malignant CBs in pathology	3.41 (3.16 to 3.66)
Time to process non-malignant CBs in pathology	2.78 (2.82 to 3.14)
Time to process LNBs in pathology	3.38 (2.93 to 3.83)
Time to process surgical specimens in pathology	9.64 (8.91 to 10.37)
Time to report ER test on CB	5.60 (5.14 to 6.07)
Time to report ER test on surgical specimen	13.42 (10.48 to 16.35)
Time to report IHC test on CB	6.49 (5.92 to 7.01)
Time to report IHC test on surgical specimen	17 (13.53 to 20.47)
Time to receive FISH test result	11.58 (9.06 to 14.10)
Time to first surgery from MDT	18.32 (16.70 to 19.94)
Time to second surgery from MDT	14.21 (11.61 to 16.82)
Treatment period: Neoadjuvant chemotherapy and Herceptin	305 days
Treatment period: Neoadjuvant Herceptin	90 days

In addition to the time estimates input into the model at the relevant activities, a couple of other time parameters were considered to complete the model in Simul8.

1. Arrival rate

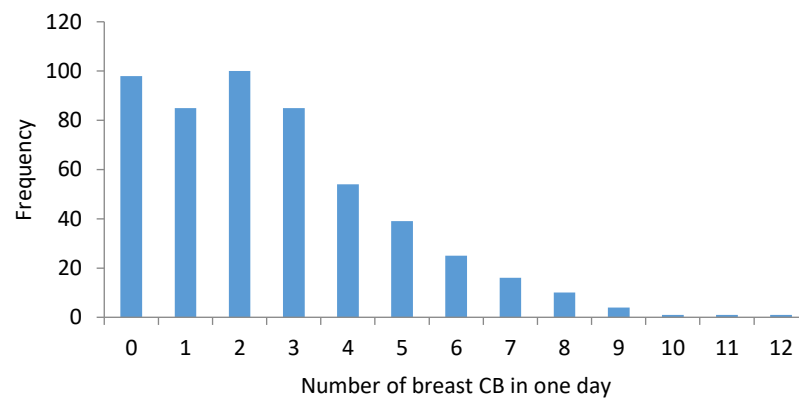
On any given day, the number of breast cores arriving at the pathology lab at UHCW ranged from 0 to 12 cores with a mean of 2.36 arrivals per day. A summary of the data is given in figure 5-5.

It was not possible to record the exact point in the day of arrival for each specimen and so for the model an inter-arrival time was calculated by taking the average number of biopsies arriving in one day and spacing these equally through the day using the following equation:

$$\text{Inter-Arrival Time} = \frac{1}{\text{Arrival Rate}}$$

The inter-arrival time was worked out to be 0.381 days. This was used in the model at the arrival activity. It was set to an exponential distribution.

Figure 5-5 Frequency of breast core arrivals at UHCW



2. MDT length

The MDT activity in the DES model is based on UHCW schedules where meetings take place twice a week on Tuesdays from 12:45 to 13:30 and on Thursdays from 12:30 to 14:00. The length of time each breast case spends at this activity is equal to the time required for MDT to discuss each case as reported in the National Cancer Peer Review Self-Assessment report for UHCW. 15 cases are discussed at Tuesday's meeting and 40 cases are discussed at Thursday's meeting and so the MDT activity's timing is dependent on the day of the week. For Tuesday, it will take on average 3 minutes to discuss each case and on Thursdays on average, it will be 2.25 minutes per case. A review of MDT effectiveness across several specialists by the London Cancer Network found that the average discussion time per case was 3.4 minutes, ranging from 1.8 to 8.6 minutes (Mughal and Goodman, 2017). In addition, a clinical audit of specifically breast MDTs found the time for discussion per case to be an average 3 minutes ranging from 2 to 4 minutes (Ruiz-Casado et al., 2014). UHCW's time for

discussion of each case falls within the ranges expressed in both studies and so it is appropriate to use these time values in the simulation.

The model is setup so that the MDT activity only accepts breast cases to travel through during the set times on a Tuesday and Thursday. Outside these times the breast cases travelling through the model wait in a queue preceding the activity.

3. Simul8 clock

The DES model in Simul8 has an inbuilt clock that controls the times the model is running and how long the model runs for. The Simul8 clock properties were set based on UHCW working practices. Histopathology labs run from 7am-6pm Monday to Friday. This is the department open the longest hours in the pathway so the clock is set to run for this time. All travel time between activities in the model are set to zero under clock properties and added manually. Results are collected and analysed for a period of one year; 52 weeks starting at 7am on a Monday and ending on Friday at 6pm. Work items arrive into the model for a period of 1 year but the model is run for 3 years to ensure all work items have completed their pathways when the model stops. This is done by setting a results collection period of 780 days but adding a constraint of 260 days at the arrival point limiting the duration of arriving work items to one year.

Cost data

In addition to time distributions, costs were assigned to all the activities in the model, including pathology activities, surgery, MDT meetings and systemic treatment. Costs were based on 2017 estimates and were obtained from various sources. These included NHS reference costs (2016-2017), BNF 73 (March 2017), UHCW costs and published literature where unit cost values could not be sourced from the other references. The model takes an NHS perspective using 2016-2017 costs; the model cost parameters and their sources are given in table 5-8.

Table 5-8 Model cost parameters

Cost parameter	Cost (£)	Description	Source
Pathology process	£98.03	Cost per specimen processed in pathology	UHCW

FISH test	£115	Additional cost for outsourcing FISH	UHCW
MDT meeting	£95 (£60 - £108)	CMDT_B - Breast Cancer MDT Meetings (Cost per case discussed)	NHS Ref 2016 - 2017
WLE surgery	£967.50	JA43B - Unilateral Intermediate Breast Procedures with CC Score 0-2	NHS Ref 2016 - 2017
Mastectomy surgery	£2419.86	JA20F - Unilateral Major Breast Procedures with CC Score 0-2	NHS Ref 2016 - 2017
WLE and SLNB	£1383.95	Cost of WLE x 1.43	NHS Ref 2016 - 2017 and Pandharipande et al. (2008)
Mastectomy and SLNB	£3460.40	Cost of Mastectomy x 1.43	NHS Ref 2016 - 2017 and Pandharipande et al. (2008)
WLE and ALND	£1615.73	Cost of WLE X 1.67	NHS Ref 2016 - 2017 and Pandharipande et al. (2008)
Mastectomy and ALND	£4041.17	Cost of Mastectomy x 1.67	NHS Ref 2016 - 2017 and Pandharipande et al. (2008)
Re-biopsy	£293.64	YJ02Z - Unilateral Core Needle Biopsy of Lesion of Breast	NHS Ref 2016 - 2017
Chemotherapy	£3026.51	Average of the 4 FEC regimens	Joint Formulary Committee (2017)
Chemotherapy and Herceptin	£27626.51	£24,600 + £3026.51	NICE (2006b)+Joint

			Formulary Committee (2017)
Cavity shave	£354.47	JA45Z -Unilateral Minor Breast Procedures	NHS Ref 2016 - 2017
ALND	£1420.90	WH54B - Procedures on the Lymphatic System with CC Score 0	NHS Ref 2016 - 2017
Margin and ALND	£1775.37	JA45Z + WH54B	NHS Ref 2016 - 2017
Hormone therapy	£790.83	5 year course of Tamoxifen	Joint Formulary Committee (2017)

The 2017 pathology process costs per specimen were obtained from the finance manager at UHCW. £98.03 is the cost of handling each specimen in the pathology lab regardless of the number of sections. This is a one off cost encompassing all specimen processing activities and tests carried out onsite, for specimens that have an ER and an IHC test also, the costs of these tests are included in the £98.03. For this reason all activities in the model associated with ER and IHC HER2 testing do not have a cost assigned.

The model follows UHCW processes where FISH is outsourced and as such this is reflected in an extra cost at FISH activities. A total cost of £115 is assigned to all FISH related activities; this is based on UHCW values and is made up of £110 for tests at the external laboratory and £5 delivery costs.

No reference costs are available for breast surgery carried out alongside a SLNB or ALND; this was also reported by Cooper et al. (2011), Huxley et al. (2015) and Rafia et al. (2016). All three studies adopted the same method to calculate the costs of these procedures; the same method was adopted here. Pandharipande et al. (2008), a US based economic evaluation reported the cost of breast surgery alone, with SLNB and with ALND. The ratio of the cost of breast surgery alone to that with each of the lymph node interventions was calculated. This ratio was then used with the NHS reference costs for WLE and Mastectomy to estimate the costs to be included in the model at WLE with SLNB or ALND and Mastectomy with SLNB or ALND.

Chemotherapy is given alone or in combination with Herceptin depending on the patients' grade and HER2 status. The most commonly prescribed chemotherapy at UHCW is the FEC

group but at different doses and with or without docetaxel. Table 5-9 gives an overview of the various FEC regimens given at UHCW. The model assumes patients are given 6 cycles of FEC or 3 cycles of FEC and 3 cycles of T when docetaxel is prescribed. The cost for each regimen is worked out based on 1.71m^2 ; the average body surface area of a female cancer patient in the UK (Sacco et al., 2010). The total cost calculation assumes any unused content in vials is discarded. The cost for chemotherapy used in the model and reported in table 5-8 is the average of the four regimens in table 5-9 below.

Table 5-9 Summary of chemotherapy regimens at UHCW

	FEC 75	FEC 100	FEC-T 75	FEC-T 100
Dose	Fluorouracil ($600\text{mg}/\text{m}^2$)mg	Fluorouracil ($500\text{mg}/\text{m}^2$)mg	Fluorouracil ($600\text{mg}/\text{m}^2$)mg	Fluorouracil ($500\text{mg}/\text{m}^2$)mg
	Epirubicin ($75\text{mg}/\text{m}^2$)mg	Epirubicin ($100\text{mg}/\text{m}^2$)mg	Epirubicin ($75\text{mg}/\text{m}^2$)mg	Epirubicin ($100\text{mg}/\text{m}^2$)mg
	Cyclophosphomide ($600\text{mg}/\text{m}^2$)mg	Cyclophosphomide ($500\text{mg}/\text{m}^2$)mg	Cyclophosphomide ($600\text{mg}/\text{m}^2$)mg	Cyclophosphomide ($500\text{mg}/\text{m}^2$)mg
			Docetaxel (T) ($75\text{mg}/\text{m}^2$)mg	Docetaxel (T) ($100\text{mg}/\text{m}^2$)mg
Regimen	6 Cycles FEC	6 Cycles FEC	3 Cycles FEC + 3 Cycles T	3 Cycles FEC + 3 Cycles T
Total Cost*	£2482.02	£2379.66	£2845.26	£4399.08

*F: 2.5g/100ml solution for infusion £32, E: 200mg/100ml solution for infusion £347.55, C: 1g powder solution for injection £17.06, T: 80mg/8ml solution for infusion £534.75 (Joint Formulary Committee, 2017)

UHCW patients are prescribed the 3weekly course of Herceptin for a period of 12months. NICE (2006b) Technology Appraisal (TA) works out the average cost per patient for this regimen to be £24,600. The TA is based on 2008 drug costs for Herceptin, however BNF 73 (Joint Formulary Committee, 2017, pg. 796) states the cost per 150mg vial to remain at £407.40; identical to the value used in the TA. For this reason, the £24,600 is not adapted and the same cost for Herceptin is used in this model. The cost for hormone therapy is based on a 5-year course of Tamoxifen (Joint Formulary Committee, 2017).

A DES model representing the breast cancer pathway that can be used in the evaluation of digital pathology has been developed and validated. The model parameters currently represent current practice, conventional microscope time and test result data based on UHCW data was used to inform the model parameters. This is ready for experimentation

through the modification of the appropriate parameters to reflect digitisation of pathology systems. Outputs from the DES model can be observed and measured to compare the time and cost implications for breast cancer patients prior and post-digitisation.

5.3.2 Economic evaluation

The DES model follows women with suspected early breast cancer through the stages of their management in secondary care.

Entities arriving in the DES model can be defined as all new suspected breast cancer cases with one core biopsy from which a definitive diagnosis can be made. These can be with or without a lymph node biopsy specimen attached

The model starts with the arrival of the first breast core biopsy specimen at the pathology lab and ends when the final steps of patients' treatment plans are initiated. The model does not represent recurrent or malignant cases that have spread to other parts of the body, where cancer is diagnosed these are all early breast cancers that have not been previously treated.

All cost and time model parameters are described and methods for estimation are given in the previous section, at this point the model reflects the use of conventional microscopes in pathology. The appropriate model parameters reflecting test results and time estimates that will be influenced by pathology digitisation can be manipulated to evaluate the changes in treatment prescribed and time to treatment that are expected with digitisation.

The breast cancer pathway activities that are expected to be influenced by pathology digitisation were identified, these are referred to as the experimental factors and are listed in figure 5-6. The majority of these are associated with pathology TAT (turnaround time) or the biomarker results required to guide breast cancer treatment. Furthermore, changes in MDT efficiency with DP systems were also investigated. The appropriate parameters were adjusted to reflect digitisation as described in the published literature in chapter 3 or through discussion with clinicians at UHCW; these are highlighted and explained through the remainder of this section.

5.3.2.1 Turnaround time

Recent research suggests improvements in the time to diagnosis with digitisation. One study; Vodovnik (2016) measured and reported a 6% time reduction over the diagnostic period as well as suggesting based on observations, albeit not measuring these, a 10% overall time saving in the pathology workflow when time savings in the non-diagnostic period are considered. The DES model does not differentiate between the diagnostic and non-

diagnostic elements, specimen processing in pathology is characterised by one activity encompassing both phases. The time distributions attached to these represent the time from the arrival of the specimen in pathology to the point a report is generated by the pathologist and is ready for discussion at MDT. This structural decision was largely driven by the UHCW data, independent time data for each phase of the pathology workflow was not available. With limited information, it is difficult to define the overall efficiency implications of digitisation for the pathology workflow with certainty. A 10% hypothetical time saving was applied to all the corresponding activities in the model listed with TAT as an experimental factor in figure 5-6. This was not applied to FISH testing activities as these are carried out offsite and will not be influenced by digitisation at UHCW. The time estimates in the model that reflect CM were altered in parallel to reflect DP and are used to investigate the implications of this hypothetical time saving for the breast cancer pathway.

In addition to overall time savings, the UHCW digital pathology business case suggests DP systems will make possible the availability of pathology reports for 50-60% of CBs on the same day as specimen arrival in the lab. The one day target is applied to CB processing and biomarker reporting on these specimens, depicting a second scenario for TAT.

Through discussion with UHCW pathologists it was also understood that theoretically it was possible for CBs to be turned around on the same day, within 6 hours of specimen arrival in the lab, a third scenario for TAT will investigate the implications of this hypothesis for the breast cancer pathway.

The consequences of digitisation for FISH reporting is the final TAT to be explored. Currently glass slides are sent from UHCW to an external pathology lab for FISH testing as UHCW do not have the infrastructure in place to perform this test onsite. The benefits of digitisation will be realised for this particular biomarker test only if the reporting laboratory also has digital systems set up to accept and examine electronic slides, in this situation glass slide delivery times will be eliminated. For FISH TAT, under the digital pathology scenarios, the model assumes that the reporting lab also has access to digital systems and the time to obtain a FISH result is equal to the time to report on IHC since delivery times are reduced from days to seconds.

Figure 5-6 Experimental factors

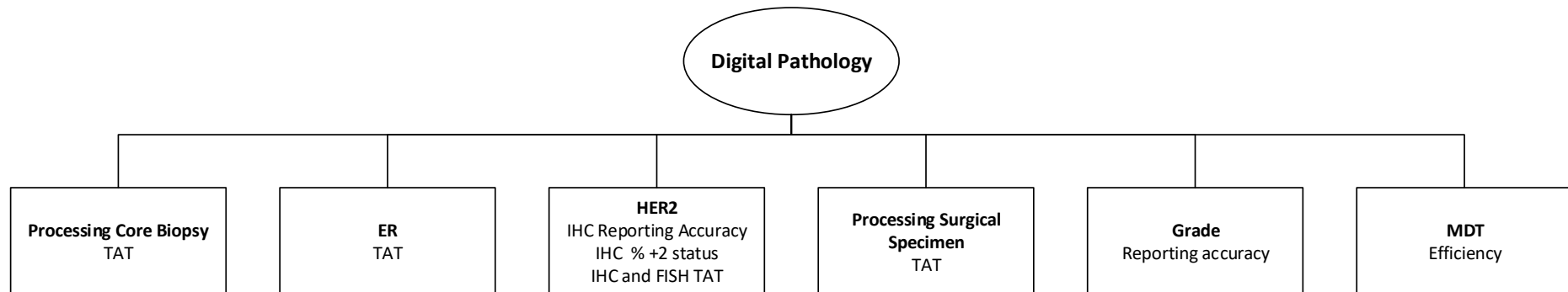


Table 5-10 Summary impacts of Digital Pathology

Time implications			Accuracy implications		
Parameter	Digital Pathology	Reference	Parameter	Digital Pathology	Reference
Pathology TAT	10% time saving in pathology activities	Vodovnik (2016)	Grade results	If HER2 negative: Grade1-29%	Longacre et al. (2006)
	50% of CB cases reported in 1 day	UHCW Business Case		Grade2-43%	
	100% of CB cases reported in 6 hours	UHCW pathologists		Grade3-28%	
MDT	All cases discussed in 2minutes	Ruiz-Casado et al. (2014)		If HER2 positive: Grade2-32%	
FISH TAT	Equal to IHC TAT	----	HER2 IHC results	22% reduction in 2+ scores 7% reduction in positive cases 9% increase in negative cases	(Dobson et al., 2010)

5.3.2.2 MDT efficiency

The current time taken to discuss each case at UHCW MDT ranges from 2.25 to 3 minutes depending on the day of the week. Research into DP impacts on MDT are limited and only one study truly measured and compared the MDT discussion time per case when conventional and digital systems are in place. Thorstenson et al. (2014) observed a 40% decrease in discussion time, from 6 minutes with conventional to 3.6 with digitisation per case. Currently without DP, UHCW cases take less time than this to discuss at MDT, a 40% reduction in current discussion time will imply cases are discussed in 1.35 or 1.8 minutes depending on the day of the week which logically is not practical.

A second study, Ruiz-Casado et al. (2014) implemented an audit of cancer MDT meetings, they reported an average 3 (2 to 4) minute discussion time per breast case which falls in line with UHCW data. Although Ruiz-Casado et al. (2014) did not particularly investigate digitisation, under the digital pathology scenario the effects on the breast cancer pathway if all cases are discussed by MDT in 2 minutes i.e. the lower bound of time when conventional microscopes are used for pathology reporting, was explored.

5.3.2.3 Reporting Accuracy

In addition to the pathology workflow and MDT impacts, digitisation of pathology systems are expected to have influence on the accuracy of test results as described in chapter 3, particularly where image analysis alongside digital pathology systems are used to supplement pathologist's reporting. Notably for breast cancer patients the outcomes of interest are ER, HER2 and grade results.

ER

It was evident through discussion with pathologists at UHCW that they expected minimal or no impact of digitisation on the ER status reported. They already considered these results to be fairly accurate. The ER biomarker test is generally easy and quick to carry out, it follows a binomial scoring pattern, tumours are either strongly positive or negative, very rarely will there be a mid-range reading that is difficult to interpret. Among the digital pathology research described in chapter 3 there were no published study reporting the accuracy of ER scoring with each of the two pathology techniques but several measured and reported a good level of concordance at best showing ER scoring with digital to be equivalent to that with conventional microscopes. The economic analysis assumes ER result proportions are not influenced by pathology digitisation, the model distribution representing this test result that is used for the conventional analysis remains identical under the digital model scenarios.

HER2 status

The second breast cancer test to be considered is that to determine HER2 IHC status. Dobson et al. (2010) published study specifically compares HER2 IHC scoring using conventional microscopes to digital pathology with image analysis and measures these against an agreed reference standard to define the reporting accuracy of each approach for this biomarker test. Sensitivity and specificity are improved with digitisation with the former increasing from 65% to 68% and the latter from 99% to 100%. Furthermore the share of equivocal results reported fall from 23.5% with conventional to 18.4% with digital.

UHCW data showed the divide between the three HER2 IHC scores to be 14% as positive, 23% as equivocal and 63% as negative results. If the result changes with digitisation observed by Dobson et al. (2010) are applied to the UHCW data, the HER2 results will split so that 13% of all cases are positive, 18% are equivocal and 69% are negative. A model distribution to reflect these results is generated and input in the model at the appropriate parameters when conducting the digital pathology analysis.

Grade

The accuracy of the grade result reported on the surgical specimen is the final model parameter that is predicted to be influenced by digitisation. As with ER results, no published study has carried out a head-to-head study comparing grade using CM and DP. Longacre et al. (2006) studied the accuracy of grade classification and reported this for each of grades 1, 2 and 3 breast cancers. They compared the breast cancer grade reported using conventional microscopes to an agreed reference standard. The results of their study are given in table 5-11.

Table 5-11 Accuracy of breast cancer grade reporting

Grade	Mean	Lower	Upper
1	83.3%	75%	100%
2	64.6%	50%	83.3%
3	92.3%	79%	100%

Since there is no current published research on the implications of digitisation for breast cancer grade a hypothetical scenario is constructed. It is assumed that the accuracy of the grade distribution reflecting UHCW results that is input

into the model is accurate at the mean rates reported in Table 5-11. An assumption is made that with digitisation, accuracy improves to the upper limits reported in table 5-11 so that 16.7% of grade 1 cancers are upgraded with DP and 7.75% of grade 3 cancers are downgraded. For grade 2 cancers, HER2 status is also considered alongside since the model includes a separate grade distribution for each HER2 result. 18.7% of grade 2 cancers are

assumed to alter with DP. For the HER2 negative cases the 18.7% error rate is split equally i.e. 9.4% are downgraded to grade 1 and 9.4% are upgraded to grade 3. For the HER2 positive group there are no grade 1 cancers so an assumption is made that all these are upgraded with digitisation to grade 3 cancers.

5.3.2.4 Scenario Analysis

Due to the extent of the impact of digitisation there were several model parameters that need to be altered to reflect this scenario. These include both time and test result parameters as described in table 5-12. The scenarios constructed for experimentation are all hypothetical due to the limitations of the primary research around the impacts of digitisation.

Table 5-12 Experimental scenarios

CM: Base case scenario				
Pathology TAT		As described in table 5-7		
MDT		Tuesday: 3min/case Thursday: 2.25min/case		
FISH TAT		11.58 days		
Grade	If HER2 negative	Grade 1: 30%	Grade 2: 44%	Grade 3: 26%
	If HER2 positive	Grade 1: 0%	Grade 2: 33%	Grade 3: 67%
HER2 IHC results		Negative: 63% Equivocal: 23% Positive: 14%		
DP: Scenario 1				
Pathology TAT		10% overall time saving at all pathology related activities		
MDT		All breast cases discussed in 2min		
FISH TAT		Equal to IHC TAT		
Grade	If HER2 negative	Grade 1: 29%	Grade 2: 43%	Grade 3: 28%
	If HER2 positive	Grade 1: 0%	Grade 2: 32%	Grade 3: 68%
HER2 IHC results		Negative: 69% Equivocal: 18% Positive: 13%		
DP: Scenario 2				
Pathology TAT		50% of CB cases reported in 1 day		
MDT		All breast cases discussed in 2min		
FISH TAT		Equal to IHC TAT		
Grade	If HER2 negative	Grade 1: 29%	Grade 2: 43%	Grade 3: 28%
	If HER2 positive	Grade 1: 0%	Grade 2: 32%	Grade 3: 68%
HER2 IHC results		Negative: 69% Equivocal: 18% Positive: 13%		

DP: Scenario 3				
Pathology TAT		100% of CB cases reported in 6 hours		
MDT		All breast cases discussed in 2min		
FISH TAT		Equal to IHC TAT		
Grade	If HER2 negative	Grade 1: 29%	Grade 2: 43%	Grade 3: 28%
	If HER2 positive	Grade 1: 0%	Grade 2: 32%	Grade 3: 68%
HER2 IHC results		Negative: 69% Equivocal: 18% Positive: 13%		

In order to grasp the full extent of potential implications, three hypothetical scenarios were drawn up as listed in Table 5-12. The first explores the impacts of an overall 10% time saving in workflow for all pathology related activities, the second investigates implications if CBs are reported on in one day from arrival in the lab and the third if CBs are reported on in 6 hours. All three scenarios are altered similarly to reflect test results, FISH TAT and MDT efficiency when digital methods are used in pathology.

5.3.2.5 Output Analysis

Four model outputs were measured under each digital pathology scenario and compared against the CM model outputs. These are:

1. The time each breast patient spends in the pathway

The time effect of pathology digitisation for breast cancer patients is one of the outcomes to be investigated. The DES model allows the tracking of different groups of patients through the breast cancer pathway recording the time each spent in the DES model. All breast cases have the same start point i.e. the arrival of CBs in pathology. However each breast case will take a different route through the model each with its own endpoint depending on the initial CB result. The total time spent in the model pathway will vary, therefore the model time outputs were analysed separately dependent on the initial CB result of the breast case. Patients can be split into 4 groups as described in table 5-13.

Table 5-13 Patient groups to be analysed independently

Initial CB result	Arrival Point	End Point
Benign (B1/B2)	Arrival of CB with definite diagnosis in pathology	Date of discharge
In situ (B5a)		Date of discharge
Invasive (B5b) <i>Neoadjuvant</i>		Date of oncology meeting post-surgery where final systemic treatment agreed with patient

Invasive (B5b) <i>Adjuvant</i>		Date of oncology meeting where adjuvant treatment plan agreed with patient
-----------------------------------	--	--

2. The recommended treatment packages

In addition to time savings, accuracy adjustments resulting from pathology digitisation will impact the treatment prescribed. Simul8 allows the capture of this information by counting the number of work items passing through an activity related to a treatment package, facilitating the comparison of treatment plans before and after digitisation.

3. The number of FISH tests requested

Accuracy adjustments in particular those related to HER2 assessment whereby the probability of reaching a definitive diagnosis on the first IHC test improves will also impact the number of times cases are discussed at MDT and the number of external FISH tests ordered.

4. The number of MDT discussions per patient

5. Cost/patient

All performance measures and the simulation object from which they are collected are given in table 5-14. The inbuilt Simul8 trial calculator was used to determine the number of runs to be carried out to reach a 5% precision level for each performance measure. The results are given in table 5-14 and based on this the model is run 1114 times for each scenario trial.

Table 5-14 Simul8 trial calculator results

Outcome	Simulation Object	Performance Measure	Recommended Runs for 5% precision
Treatment plans	Neoadjuvant chemotherapy	Number completed Jobs	334
	Adjuvant chemotherapy	Number completed jobs	12
	Neoadjuvant chemotherapy and Herceptin	Number completed Jobs	730

	Adjuvant chemotherapy and Herceptin	Number completed jobs	26
FISH test ordered	Three FISH related activities included in the model	Number completed jobs	30
		Number completed jobs	33
		Number completed jobs	1114
MDT discussions	MDT	Number completed jobs	9
Time to endpoint	Benign cases	Average Time in system	4
	DCIS cases	Average time in system	4
	IBC and neoadjuvant	Average time in system	125
	Invasive and adjuvant	Average time in system	4

The original version of the model representing the use of conventional microscopes in the pathology lab was initially run to generate outputs for conventional microscope for each of the performance measures against which the digital pathology outputs from each of the three scenarios can be compared. The pathway is modified to reflect each of the scenarios listed in table 5-12, outputs are generated and compared to those of the conventional microscope scenario. Due to the many hypothetical impacts of digitisation, the inclusion of many experimental scenarios will identify the model parameters that are likely to have the most impact and need to be explored further to ascertain their influences on the breast cancer pathway.

The economic evaluation takes an NHS perspective and follows patients through secondary care. A one-year time horizon over which costs and outcomes are evaluated is sufficient and discounting does not need to be applied.

The arrival rate of breast core biopsy specimens in the pathology lab does not change with digitisation. The mean number of breast cases arriving in the model over a one-year period is 682, the model has been programmed to follow these through to the various end points depending on the initial CB result as described earlier. After running for one year, the model prevents any further breast cases arriving in the model but continues to allow already entered cases to follow through and complete their pathways. The model is designed to stop running once all breast cases have reached their appropriate endpoints

5.4 Results

5.4.1 Model validation

Prior to carrying out any analysis, the model was validated by running it and comparing it to the real system, the breast cancer pathway at UHCW. If the inputs to the model are the same as reality then the outputs from the model should also be reasonably similar to real data. The total time a breast cancer patient spends in the model is compared to the data collected in order to validate the model. However as there are two groups of cancer patients dependent on their treatment plans and each with a different end point these are first separated out.

The first model output to be compared is the time it takes for a breast case to reach the point of adjuvant treatment planning in oncology from CB specimen arrival in pathology. In particular these are cases that have been diagnosed as invasive whose treatment plan was made up of surgery followed by adjuvant therapy. The data collected from UHCW is compared to the model output. Figure 5-7 illustrates the point estimate of the number of days with 95% confidence intervals and figure 5-8 shows the results of a two sample t-test. At the 0.05 level of significance there is no evidence to suggest that there is a difference in the mean total time between the model output and the data collected at UHCW.

Figure 5-7 Days from CB arrival in pathology to adjuvant treatment plan

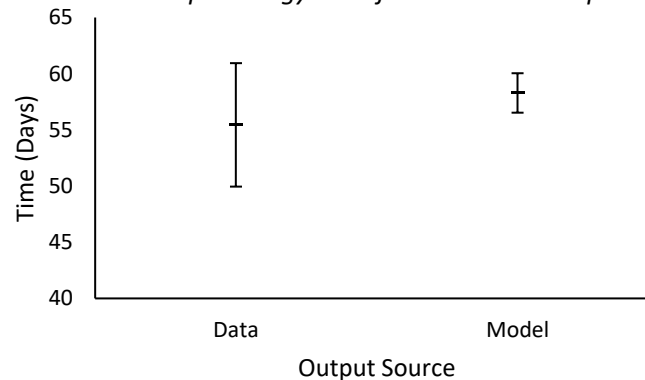


Figure 5-8 Two sample t-test comparing data and model output

Two-sample t test with equal variances

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
Data	45	55.44444	2.726146	18.28755	49.95026	60.93863
Model	236	58.28546	.8914004	13.69395	56.5293	60.04162
combined	281	57.83049	.8666003	14.52687	56.12461	59.53637
diff		-2.841014	2.361108		-7.488863	1.806835

diff = mean(Data) - mean(Model) t = -1.2033
 Ho: diff = 0 degrees of freedom = 279

Ha: diff < 0
 Pr(T < t) = 0.1149

Ha: diff != 0
 Pr(|T| > |t|) = 0.2299

Ha: diff > 0
 Pr(T > t) = 0.8851

Identical steps were also carried out for the patients recommended neoadjuvant therapy in the dataset and in the model. For this group the time from CB arrival in pathology to discharge from secondary care is calculated. The mean output from the model at this stage is 227 (95%CI: 125, 330) days with the time for neoadjuvant cases in the dataset falling within the model's confidence interval.

5.4.2 Conventional microscope outputs

The first stage in determining the impacts of digitisation is to understand the model outputs when conventional microscopes are used in the pathology lab. The model was run as it is constructed and populated initially, reflecting current standard practice for breast cancer. The outputs for this scenario against which those from the three digital pathology scenarios were compared are given in table 5-15.

The number of completed jobs recorded at the MDT activity includes all suspected breast cancer cases referred for investigation. Patients with a benign or in situ diagnosis are relatively simple to handle and the number of MDTs they are discussed at will not be influenced by pathology digitisation since no supplementary testing beyond the basic tissue analysis is carried out on these specimens. Contrarily, invasive cases are a lot more complex to manage and will present at MDT meetings at least twice for discussion. These are separated from the remainder of breast cases and the number of times each patient is discussed at MDT is recorded. Each invasive breast cancer case is discussed on average 4.5 times when conventional microscopes are used, with nearly 60% of patients presenting at 3 or 4 MDT meetings over the course of their breast cancer pathway. Summary of the number of MDT discussions per invasive breast cancer case is given in figure 5-9.

Figure 5-9 Histogram of the number of times an invasive breast cancer case is discussed at the MDT meeting

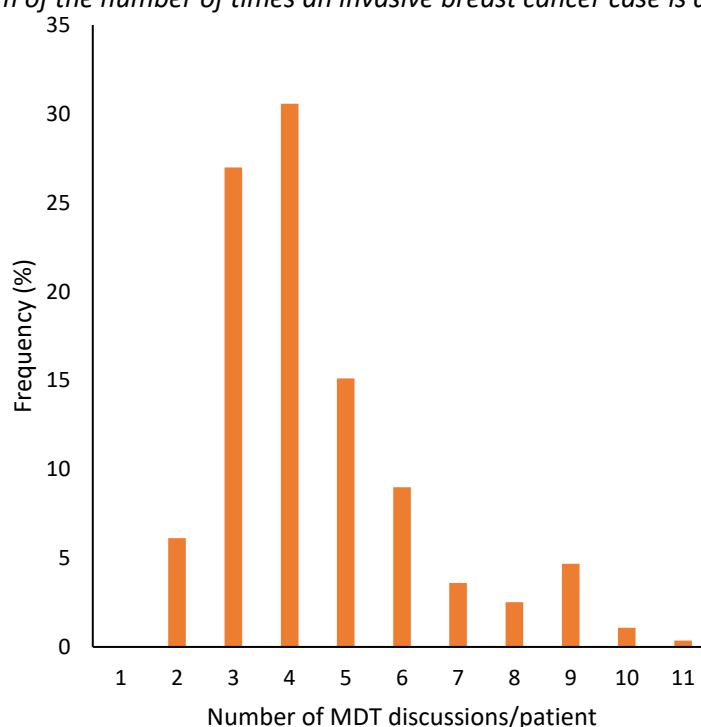


Table 5-15 Model outputs under each scenario

Model output	CM	DP1	DP2	DP3
Average time in pathway (days)				
Benign cases	4.46	3.90	3.21	1.76
	Δ	↓0.56	↓1.25	↓2.7
DCIS cases	34.40	32.45	32.69	31.08
	Δ	↓1.95	↓1.71	↓3.32
IBC cases: Neoadjuvant treatment	251.53	231.62	233.24	240.33
	Δ	↓19.91	↓18.29	↓11.2
IBC cases: Adjuvant treatment	60.22	55.81	55.42	52.52
	Δ	↓4.41	↓4.8	↓7.7
Annual number of patients recommended each treatment package				
Neoadjuvant: Chemotherapy	5.73	5.95		
Neoadjuvant: Chemotherapy and Herceptin	2.60	2.18		
Adjuvant: Chemotherapy	123.74	132.14		
Adjuvant: Chemotherapy and Herceptin	65.39	55.98		
Annual number of FISH tests requested	98.14	76.23		
	Δ	↓21.91		
Annual number of MDT discussions	1616.12	1496.99	1366.98	1248.76

	Δ	↓119.13	↓249.14	↓367.36
Annual total cost of systematic therapy				
Neoadjuvant: Chemotherapy	£17,354.89	£17,996.05		
Neoadjuvant: Chemotherapy and Herceptin	£71,819.01	£60,163.30		
Adjuvant: Chemotherapy	£374,510.24	£399,934		
Adjuvant: Chemotherapy and Herceptin	£1,806,461.28	£1,546,489		
Total systemic therapy cost	£2,270,145.42	£2,024,582.35		
	Δ	↓£245,563		
Annual cost of FISH testing	£11,286.62	£8,765.70		
	Δ	↓£2520.92		
Annual cost of MDT	£153,531.08	£142,213	£129,863	£118,631
	Δ	↓£11,317	↓£12,667	↓£34,889
Annual Δ costs				
Δ Total		-£259,401	-£271,751	-£282,983
Δ Cost /patient		-£380	-£398	-£414
Δ Cost /IBC patient		-£1085	-£1137	-£1184

5.4.3 Digital pathology outputs

The model outputs are not influenced by all the experimental factors. The treatment plans prescribed and the number of FISH tests ordered are driven by test results, which do not change across the three digital pathology scenarios, one model output distribution is given for each of these performance measures. The number of MDT discussions and the time to end-point for each breast case type are influenced by both the accuracy and time experimental factors, for these measures an output distribution is given for each digital pathology scenario. The results of the digital pathology scenarios are given in table 5-15 alongside the CM outputs.

The DP scenario model outputs suggest changes in breast cancer treatment plans with digitisation; these are very small for the neoadjuvant cases primarily due to the very limited sample size. However for adjuvant cases the variations are greater, ten patients will avoid chemotherapy treatment altogether and 9 patients are downgraded from Herceptin and chemotherapy regimen to chemotherapy treatment only. This change corresponds to an overall cost saving since the combined treatment plan is associated with greater cost. In addition, more than one in every five FISH tests ordered under the conventional microscope scenario are avoided with digitisation, corresponding to over £2500 saved on offsite FISH testing and contributing to time savings in the pathway whilst awaiting the arrival of external test results.

Similar trends are realised for the total number of MDT discussions, these are reduced across all three digital pathology scenarios corresponding to annual cost savings of approximately £11000, £24000 and £35000. Through modelling it was possible to combine the impacts of changes in pathology workflow on MDT discussions per case with the effects of changes in accuracy, particularly the reduction of FISH testing and the increased availability of definitive HER2 results on the first round of testing. A histogram of the number of MDT discussions per case for each pathology scenario is given in figure 5-10.

DP1 scenario reduces the upper limit of MDT discussions/patient from 11 with CM to 8 presentations. Over 60% of cases are now discussed at a total of 3 or 4 meetings. DP2 reduces the upper limit to 7 days whilst maintaining that over 60% of cases are discussed at 3 or 4 meetings. DP3 reduces the maximum even further to 5 meetings and ensures that over 80% of cases are present at no more than 3 meetings for discussion. Figure 5-11 illustrates a statistically significant reduction in the mean number of MDT meetings/patient,

decreasing from an average 4.5 discussions IBC patient with CM to 4, 3.5 and 3 with DP1, DP2 and DP3 respectively.

Figure 5-10 Histogram of the number of MDT discussions/IBC patient

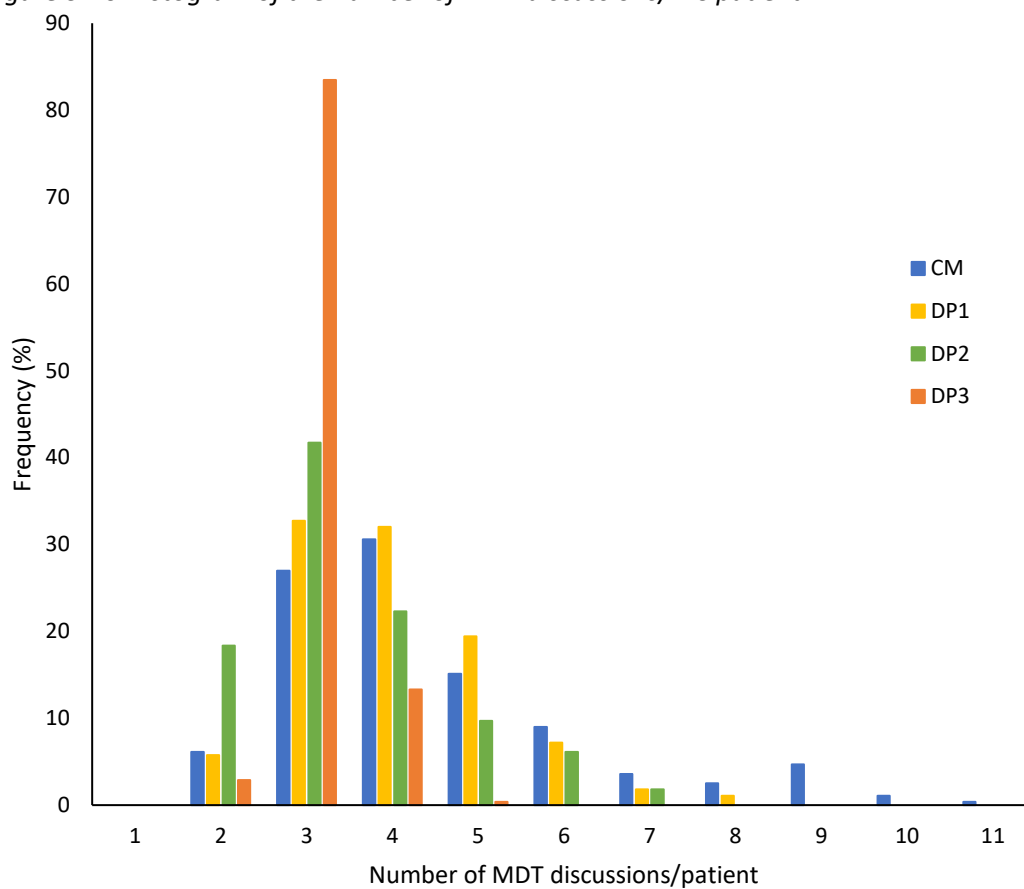
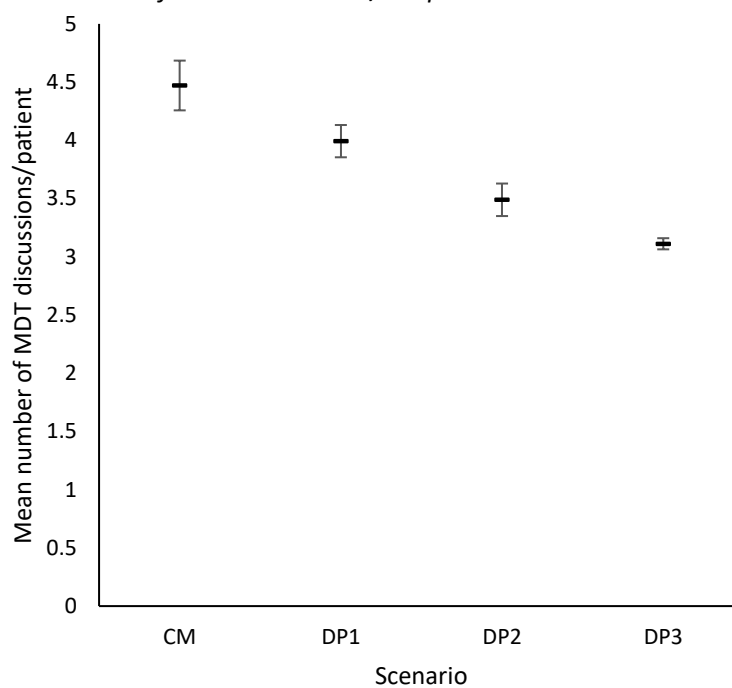


Figure 5-11 Mean number of MDT discussions/IBC patient



The final performance measure of interest is average time in the breast cancer pathway for each group of patients. Across all three DP scenarios, time savings are observed for all the types of breast cases, whether these are benign and discharged or are invasive and are exposed to the full extent of activities in the pathway. For the benign cases, time savings in the pathway ranged from half a day to over 2.5 days. For the DCIS cases this ranged from nearly 2 days to over 3 days. For the IBC cases time savings can be as high as 7.7 days if patients are recommended an adjuvant treatment plan or over 11 days where a patient is recommended a neoadjuvant treatment plan.

As with MDT discussions, modelling allows the analysis of the combined time impact of pathology workflow efficiency and accuracy effects. The changes seen in time to End for benign cases and IBC cases are due to pathology workflow improvements only. For time to END for IBC cases who receive adjuvant treatment, time savings result from a combination of pathology workflow efficiencies and test result changes. Finally for IBC cases who receive neoadjuvant treatment the time savings in the pathway are the greatest, this is largely due to the inclusion of time for neoadjuvant treatment being in the time to end-point. The time savings observed are a combination of enhancements in pathology workflow, accuracy and changes in treatment plans which influence the time taken to administer neoadjuvant therapy.

When all cost consequences were combined, cost savings were observed across all three DP scenarios. For every patient referred to secondary care with suspected breast cancer the total cost saving ranged from £380 to £415. All performance measures on which the total cost savings are based are borne as a result of changes in the pathways of IBC patients. Benign and DCIS cases do not undergo FISH testing or supplementary reporting beyond the classification of the type of cancer cells. These cases are discussed at MDT once if benign and twice when DCIS. When the cost savings are attributed to IBC cases only they ranged from £1085 to £1184 per patient depending on the DP scenario under investigation.

5.5 Discussion

A DES model representing the breast cancer pathway that can be used in the evaluation of pathology interventions has been developed and validated. The model was used to measure the costs and consequences of pathology digitisation through the manipulation of the model parameters that correspond to the experimental factors of digitisation. Outputs from the DES model were measured to compare the time and cost implications for breast cancer patients prior and post pathology digitisation.

The economic analysis showed pathology digitisation to result in time savings in the pathway for all types of suspected breast cancer cases, whether they are determined to be benign and discharged or are invasive cases that go on to surgery and systemic treatment. All breast cases experience a time saving in their pathway under all of the three scenarios that represent pathology digitisation. Furthermore, there was a reduction in the number of MDT discussions for invasive cases and the number of FISH tests requested with evidence of an overall cost saving across all the DP scenarios when compared to the use of CM.

Accuracy of test results drives the largest cost saving, producing changes in the systemic therapy prescribed, the number of FISH tests ordered and in addition playing a role in adjustments in the number of MDT discussions over the time horizon of the study. The cost of outpatient appointments are not computed in the model, these run in parallel to MDT meetings, a reduction in MDT will result in a similar decrease in OPAs, leading to greater NHS cost savings, better use of clinician time and avoiding unnecessary hospital visits by the patient. However the cost-savings listed do not take account of the investment needed to setup digital pathology systems. According to the UHCW business case digital pathology requires nearly £2,000,000 of investment over the initial 7years of the project, this is equivalent to approximately £277,000 per annum. This figure is very similar to the annual cost savings predicted under each of the breast cancer digital pathology scenarios considered. Under scenario 1 there is a £259,000 cost saving, scenario 2 there is over £270,000 saved and for scenario 3 over £282,000 is saved. Undeniably the benefits of digitisation will not stretch to the breast cancer pathway only but to other cancers and beyond these to other disease areas. If similar cost savings are observed for each of these disease areas, it will certainly make digital pathology a cost-saving venture.

The time savings observed under each DP scenario will support the realisation of government cancer waiting time targets. The 62-day target from arrival in secondary care through to the date of treatment overlaps with the start and end point in the model of invasive breast

cancer cases that are recommended surgery followed by adjuvant treatment. These cases experience a 4.4, 4.8 or 7.7 daytime saving under each of the DP scenarios. Under the CM scenario the 62 day target is already met at UHCW, the average time being 60 days with 96% of breast cases starting treatment within 62days. Under the DP scenarios this period is reduced further.

A similar trend was seen when analysing the UHCW data around the time for discussion of each case in MDT, it appeared that UHCW was well within the standard ranges for this. UHCW is a large teaching hospital with the infrastructure and resources in place to provide the best standard of care for patients. For features of the breast cancer pathway that already are at maximum efficiency at UHCW, digitisation has little influence, this may not be true for smaller, less well maintained hospitals. Digital pathology can be a means by which to get standards of care closer to those seen at a larger teaching hospital, reducing inequalities in diagnosis and treatment and improving outcomes, here the impacts of digitisation can be greater. The appropriate model parameters, especially those associated with time can be manipulated to reflect practice at smaller hospitals, the outcomes from such an analysis can be compared to the outcomes reported when time parameters are based on UHCW data to overcome the issues of generalisability of the results.

The economic analysis described in this chapter presents results for three different scenarios representing digital pathology. Whilst the CM analysis is based on UHCW patient level data and is not hypothetical, the impacts of pathology digitisation that are applied to the CM model are in most cases hypothetical. The changes in HER2 test results reported with digitisation is based on a head-to-head study, however grade is based on the hypothesis that digitisation will eliminate current levels of inaccuracy in grade reporting. No primary research has compared the breast cancer grade reported using CM and DP. Identical HER2 and grade changes are applied across the three scenarios.

While the results of the economic analysis are promising for digitisation the primary research used to inform the experimental factors are not without flaws. Research into the impacts of pathology digitisation is limited, the scenarios setup are largely based on assumptions. Ideally head-to-head studies comparing results for each of the experimental factors under the conventional and digital scenarios are needed. Based on these, distributions can be constructed and input into the model for each scenario analysis. This sort of data was only available for HER2 status reporting. Information on the accuracy of breast cancer grade reporting with digitisation is non-existent and for time savings the data is very limited. This

issue was also evident in the economic analysis carried out by Ho et al. (2014), many of their model inputs were based on author assumptions rather than good-quality primary research. Evaluation of pathology services carries with it additional challenges not present across other healthcare interventions largely due to the indirect relationship between pathology and the patient. Perhaps the results of the economic analysis are overly optimistic but it is challenging to carry out a good-quality economic evaluation based on which decisions can be made with limited available research to inform the model parameters.

The aim of the economic analysis was not to estimate the cost-effectiveness of digitisation but to provide an overview of the experimental factors and model outputs that appear to make the greatest contribution to patient care in order to justify further research to support the introduction of digitisation. The accuracy of test results used to guide treatment is the experimental factor of greatest interest, it made the largest contribution to cost savings and is expected to improve patient quality of life where improved accuracy with digitisation makes for better management of treatment. Although the analysis measured changes in treatment plans, this remains an intermediate outcome and findings need to be extrapolated to determine the consequences for patient quality of life. Invasive breast cancer cases are the group most influenced by digitisation, through both accuracy changes and time savings thus justifying the focus of future research on this group of patients. Of the biomarker tests used to guide breast cancer treatment, HER2 testing was the most explored with several research publications comparing the sensitivity and specificity of this test under each of the pathology approaches. Further research around this particular biomarker is needed to consolidate the various literature and start to build a portfolio of research to support investment in digital pathology.

In reality the DES model developed can not only be used as a tool to investigate further the impacts of pathology innovation on the breast cancer pathway but can be also be used to explore the impacts of modifications at any point of the pathway. It has been developed with pathology services in mind but the availability of the schematic of the overall breast cancer pathway in chapter 4 means the model can easily be adapted to focus the investigation on other services in the pathway.

The economic benefit of digital pathology through its speed and accuracy implications is conditional on the effect of these on the patient experience through the breast cancer pathway. Changes in test results and time savings pave the way for improved and early management of disease. These are the targeted processes that are the stepping stone

between pathology and patient care. The value of pathology digitisation can only be measured when the full sequence of events is considered. Current published literature spanning across all three domains is non-existent, research highlighting the effects of digitisation on the pathology workflow and the accuracy of biomarker tests is becoming increasingly available, yet these do not go beyond these transitional points by extrapolating outcomes to breast cancer patients.

Current knowledge around the cost benefit of digitisation is scarce, one study Ho et al. (2014) measured the cost savings based on pathologist time saved and improved interpretive accuracy. However, they do not extrapolate time savings in pathology to explore time savings for the patient's cancer pathway. The accuracy improvements they assume with digitisation are based on the removal of current interpretive errors rather than a comparison of current accuracy with estimated accuracy with digitisation. Furthermore, their accuracy investigations are based on diagnosing one subtype of breast cancer correctly and they do not go as far as considering the accuracy of the underlying biomarkers that guide treatment management. Nonetheless, their economic analysis reports cost savings of \$17.73 million over the 5 year roll out period of digital pathology.

5.6 Conclusion

The economic analysis has shown digital pathology to be associated with positive outcomes for the breast cancer pathway through both time and cost savings. This analysis acts as a preliminary study in determining the cost-effectiveness of pathology digitisation. The scarcity of research comparing conventional and digital pathology outcomes for each of the experimental factors has been highlighted. As data becomes available through head-to-head studies or even through systematic review and meta-analysis, the model can be updated through the amendment of the model distributions to reflect these and the results of the analysis updated for use by decision makers.

Chapter 6: Systematic review and meta-analysis of HER2 testing accuracy

6.1 Introduction

The previous chapters investigated the impacts of pathology digitisation as a generic process by measuring its combined effects within the breast cancer pathway. Evidence of time savings in the pathway and changes in treatment plans for breast cancer patients exist. However, model probabilities reflecting changes in time and test results with digitisation were largely hypothetical or based on single-study estimates. Of the experimental factors listed in the previous chapter, the HER2 biomarker IHC test had the largest published evidence base comparing HER2 test results reported by CM and DP.

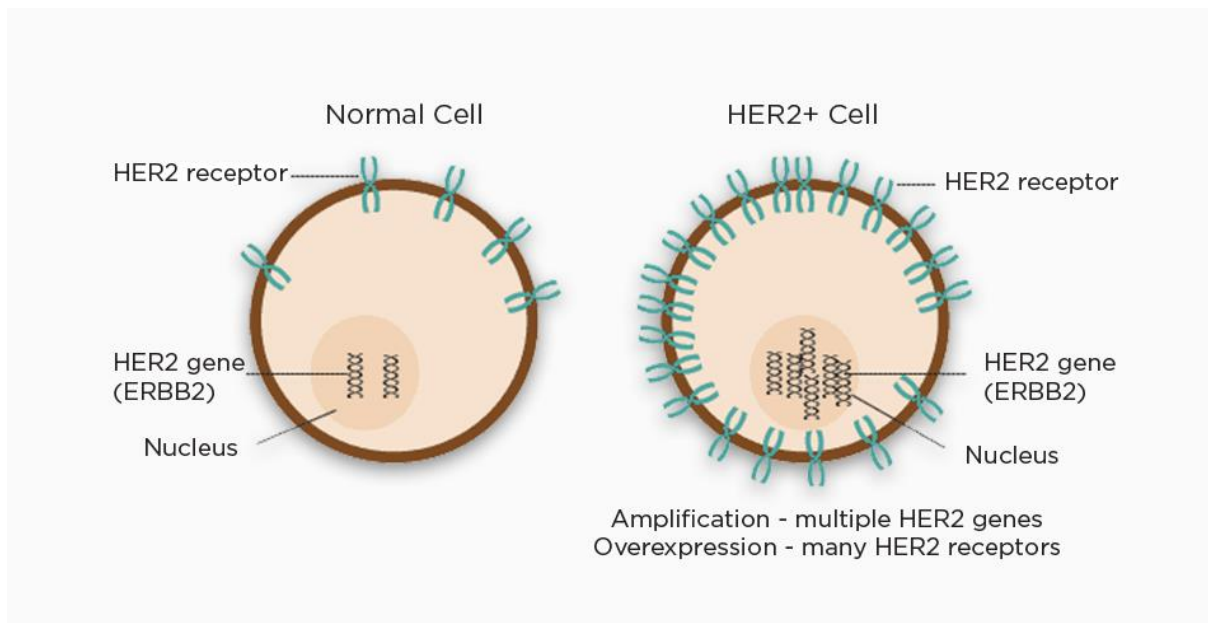
This chapter aims to build on the available literature by adopting data synthesis methods to compare the accuracy of CM and DP. Systematic review methods were used to retrieve individual studies comparing HER2 IHC scores across both arms. The accuracy data from each of the studies identified were synthesised by using meta-analysis techniques to generate pooled accuracy estimates for each of CM and DP.

6.1.1 HER2 amplification and overexpression

Human Epidermal Growth Factor Receptor 2 (HER2) is found on the surface of all cells since it is involved in normal cell growth and division (Rubin and Yarden, 2001). The HER2 receptor is encoded by a HER2 gene controlling the number of receptors on the surface of the cell. When this gene is amplified HER2 receptors are overexpressed (Slamon et al., 2001) and thus cell growth and division intensifies; a typical characteristic of cancerous cells. Figure 6-1 taken from The Journal of the Advanced Practitioner in Oncology (*JADPRO*, 2016) illustrates this process.

This situation is not found in all cancer cells as it is only one of the many ways by which they will proliferate. Breast cancers that show HER2 gene amplification or HER2 receptor overexpression have been reported to represent up to 30% or less of all cases (Mitri et al., 2012). Prior to the introduction of any targeted therapy for HER2; Slamon et al. (1987) investigated the correlation of HER2 status with relapse and survival; they found a strong and significant correlation with both. Women confirmed as positive for HER2 experienced a more aggressive form of the disease with markedly reduced disease free and overall survival (Mitri et al., 2012) compared to those with negative HER2 status. UK audits have shown that overall 14.5% of breast cancers are HER2 positive (Ellis et al., 2016).

Figure 6-1 HER2 receptor expression in normal and cancerous cells



6.1.2 Anti-HER2 therapy

Advances in medicine have resulted in the discovery of anti-Her2 therapy; specifically targeting cancerous cells that show HER2 amplification or overexpression. The monoclonal antibody Trastuzumab more commonly known as Herceptin attaches itself to the receptors on the surface of the cells blocking those signals that are telling it to continue growing and dividing (CARLSON, 2008). The clinical trial HERceptin Adjuvant (HERA) comparing Herceptin with no Herceptin showed a 50% reduction in the recurrence rate when initial results were reported at the one year point of median follow up (Piccart-Gebhart et al., 2005). Moving forward the final report on this trial further reaffirmed the improvement in disease free survival. Cameron et al. (2017) reported their findings after a median 11 years of follow up; their conclusions showed a reduction in the risk of disease-free survival (HR 0.76, 95% CI: 0.68, 0.86) and death (0.74, 95% CI: 0.64, 0.86) when one year of Herceptin treatment is given compared to just observation. With these encouraging outcomes came recommendations for Herceptin to be introduced into the treatment plans of HER2 positive breast cancer patients. NICE's first recommendations for the use of Herceptin by the NHS came in 2006 (Mayor, 2006) with the publication of guidelines for its use in early stage breast cancer (NICE, 2006b); notably this was following HERA trials publication of initial one year results.

6.1.3 Approaches to HER2 testing

It has now been established that Herceptin is an effective anti-HER2 therapy; treatment will only succeed if the cancer is confirmed to be positive for HER2 amplification or overexpression; thus it is important that true HER2 status is established correctly prior to administration of Herceptin. Data collected at UHCW has shown that all core biopsies with

an invasive result are routinely referred for HER2 testing. Any patients upgraded from a non-invasive lesion on biopsy to an invasive lesion on their surgical specimen will have their HER2 testing at this point. In most cases HER2 tests are not repeated on the surgical excision if a result is available from biopsy since research has shown high levels of concordance between both (Rakha et al., 2014). In particular a meta-analysis by Chen et al. (2012) showed over 97% specificity when HER2 results were reported on biopsy compared to surgical specimen. The specimen type used for HER2 testing does not impact results reported eliminating the introduction of uncertainty through this avenue.

6.1.3.1 IHC versus FISH

Two tests are available to determine HER2 positivity; immunohistochemistry (IHC) and fluorescence in situ hybridisation (FISH). They each measure different features of the cell to determine HER2 status. IHC counts the receptors on the surface of the cell whilst FISH quantifies gene amplification (Hicks and Schiffhauer, 2011). Though they both determine HER2 status for guiding Herceptin treatment there are differences in how they are performed, which leads to the general recommendation that IHC is carried out initially followed by FISH if results are inconclusive. HER2 IHC results on breast specimens are scored as 0 or 1+; considered negative, or 2+; showing an equivocal result or 3+; reported positive. IHC is cheaper and easier to perform than FISH (Schmidt, 2011) but reporting of results is subjective, potentially leading to inter-observer variability (Nitta et al., 2016). While this may not be a major setback for equivocal scores on IHC since they are generally confirmed by FISH, it poses a problem for 1+ and 3+ results. Here Herceptin can in theory be omitted where needed or prescribed when not required if results reported do not reflect true status since 1+ and 3+ IHC are not routinely referred for FISH testing for confirmation.

On the other hand slides for FISH are examined using a fluorescence microscope rather than a conventional bright field microscope as with IHC (Nitta et al., 2016) making it a lot more expensive and time intensive; but reporting results takes a more quantitative approach than IHC (Furrer et al., 2015). However, not all laboratories have the resources to undertake FISH testing onsite; this is especially the case at UHCW where FISH testing is outsourced not merely adding the cost of an extra test but also the further cost of delivery to and from the second laboratory. Due to the more quantitative method for scoring FISH results it is considered the gold standard in HER2 testing (Panjwani et al., 2010) irrespective of its associated shortcomings.

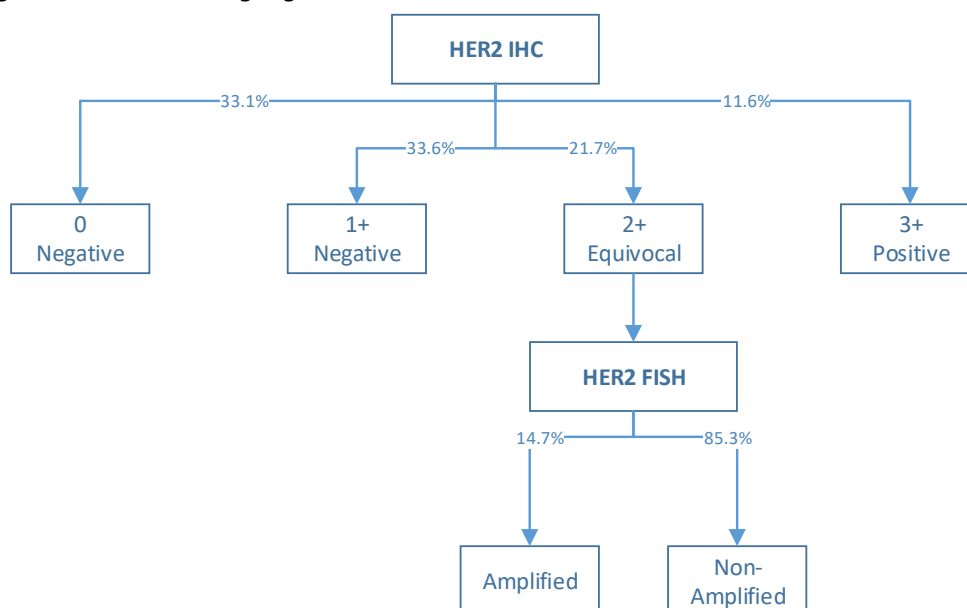
6.1.3.2 Developments in testing

Approaches for the quantification of HER2 IHC scoring; producing results closer to FISH have been suggested. Conventionally slides for IHC are prepared and viewed under a microscope by a pathologist who reports these results for discussion at MDT meetings. Digital Pathology (DP) allows these same glass slides to be scanned and viewed on a computer screen ready for examination by a pathologist. Additionally with the introduction of DP comes the opportunity for Image Analysis (IA) computer software to support the pathologist's reporting. These produce more accurate and reproducible results by way of digital measurement (Nassar et al., 2011a) rather than subjective pathologist interpretation. IA involves the use of algorithms to quantify and score a range of cell features (Webster and Dunstan, 2014) including the presence of HER2 overexpression or amplification. There are several IA software available all of which have their own integrated method for scoring HER2 on breast specimens; for ease of interpretation at MDT these are translated to the conventional IHC 0, +1, +2 or +3 score for reporting.

6.1.4 HER2 testing algorithm

Irrespective of the method that will be used for viewing IHC glass slides, HER2 testing follows a two-step process as observed at UHCW. Firstly, the IHC test is carried out followed by FISH if needed. This testing process is illustrated in Figure 6-2. The algorithm is based on UHCW testing practices which compares with current UK recommendations as described by Rakha et al. (2014) and Ellis et al. (2016).

Figure 6-2 HER2 testing algorithm



(Ellis et al., 2016)

Patients with a 2+ outcome on IHC will go on to FISH testing to confirm their HER2 status. A break-down of the proportion of the different HER2 results found during UK audits presented in Ellis et al. (2016) are given in figure 6-2 alongside the testing algorithm. 21.7% of all HER2 breast cancer cases initially had an equivocal result on IHC and required further FISH testing to verify their HER2 status (Ellis et al., 2016). A patient with a IHC 0/1+ result or non-amplified FISH will not be recommended for Herceptin whereas those with a IHC 3+ result or FISH amplified will be considered for Herceptin after taking into account other factors such as age and comorbidities.

6.1.5 Testing thresholds

Pathologists report HER2 IHC and FISH results based on recommended underlying cut-off points for each score. The American Society of Clinical Oncology/College of American Pathologists (ASCO/CAP) provide a summary of what these should be; issuing recommendations in 2007 followed by an update in 2013. The latest guidelines match those issued by the Royal College of Pathologists (RCPATH) in their most recent publication on reporting of breast disease by Ellis et al. (2016).

In order to improve the accuracy of HER2 testing; the latest ASCO/CAP guidelines modified the thresholds for scoring when compared to the previous 2007 recommendations. Table 6-1 gives an overview of the differences in these as detailed by Wolff et al. (2013).

The continual update of recommendations for scoring, ensuring they represent the true underlying HER2 status as well as giving clear and straightforward guidelines for pathologists to follow, works in the interest of both the reporting clinician and the patient due to the consequent result in accurate treatment plans. On the other hand shifting scoring thresholds over time presents a challenge when trying to summarise the diagnostic accuracy of a given test since what is determined to be a positive or negative test result has now changed. This is further amplified in the HER2 scenario as the thresholds for both IHC and FISH, which is considered the gold standard in HER2 testing have seen alterations in their thresholds between the 2007 and 2013 guidelines.

Several studies; Varga and Noske (2015), Shah et al. (2014) and Long et al. (2015) investigated the impact of the updated recommendations, with all showing similar results, an increase in overall HER2 positive cases. Thus, it is crucial any proposed summary measure of diagnostic accuracy takes into consideration the underlying method for the classification of both IHC and FISH HER2 scores.

Table 6-1 HER2 scoring thresholds, 2007 versus 2013

IHC		
Score	2007	2013
3+	Uniform intense membrane staining of > 30% of invasive tumour cell	Based on circumferential membrane staining that is complete, intense. Observed in a homogeneous and contiguous population and within >10% of the invasive tumour cells.
Equivocal	IHC 2+	IHC 2+ based on circumferential membrane staining that is incomplete and/or weak/moderate and within >10% of the invasive tumour cells; or complete and circumferential membrane staining that is intense and within ≤10% of the invasive tumour cells
1+	HER2 0: no staining HER2 1+: Weak incomplete membrane staining in any proportion of tumour cells or weak, complete membrane staining in <10% of cells	IHC 0 as defined by no staining observed or membrane staining that is incomplete and is faint/barely perceptible and within ≤10% of the invasive tumour cells IHC 1+ as defined by incomplete membrane staining that is faint/barely perceptible and within >10% of the invasive tumour cells
FISH		
Amplified	HER2 to CEP17 of > 2.2	Dual-probe HER2/CEP17 ratio ≥2.0
Equivocal	HER2/CEP17 ratio of 1.8-2.2	----
Non-Amplified	HER2/CEP17 ratio of < 1.8	Dual-probe HER2/CEP17 ratio <2.0

(Wolff et al., 2007) and (Wolff et al., 2013)

6.1.6 Accuracy of HER2 tests

Herceptin is currently only available at a high cost and is associated with cardiotoxicity reported to be as high as 27% (Keefe, 2002) . Any administration of Herceptin where it is not needed will only lead to unnecessary costs to the NHS and potential side effects to the patient with no clinical benefit. Alternatively, HER2 positive patients may lose the

opportunity of treatment if their HER2 status is not confirmed correctly; these will miss out on a potentially life changing intervention. In both HER2 positive and negative breast cancers, it is vital that results are reported as near as possible to the patient's true status to allow for the design of the most suitable care plans.

Accurate methods and clear scoring thresholds will increase pathologists' confidence in scoring HER2 especially for borderline and uncertain cases. This is particularly true for the HER2 IHC test owing to the presence of the 2+, equivocal group. These cases are referred for the more definitive FISH test. Methods increasing pathologist confidence in scoring that also reduce the number of uncertain cases will certainly shrink the total number of 2+ scores and a greater number of cases will be given a definitive result on their first round of testing. In an ideal scenario, results equivalent to FISH are reported on the initial IHC test with minimal 2+ scores. In theory this will allow the development of treatment plans most suited to the patient and will reduce the cost of testing to the NHS as well as the time to result confirmation; both as a consequence of the elimination of the need for confirmatory FISH testing.

The implications of digitisation for test reporting accuracy is one of the justifications for the use of DP IA and will form a substantial part of the rationale to support investment and the use of this technology in pathology practice.

6.1.7 Quantifying accuracy

Accuracy is an important indicator of the superiority of one testing method over another.

When summarising the accuracy of HER2 IHC there are two layers to consider:

1. Non-equivocal IHC cases that are given a +1 or +3 score
2. Equivocal IHC cases

Novel scoring methods such as DP IA, that show greater accuracy in the first group and a simultaneous reduction of cases in the second group are worthy of consideration for introduction in pathology lab procedures. To measure accuracy the results of the test of interest need to be compared against the true HER2 value, established by using FISH, the agreed upon reference standard.

6.1.7.1 Non-equivocal cases

Sensitivity and specificity are measures of diagnostic test accuracy, under the HER2 scenario, sensitivity refers to the ability of a test to detect a +3 result in patients that are HER2 positive and specificity refers to the ability of a test to detect a +1 result in patients that are HER2 negative (Lalkhen and McCluskey, 2008). The calculation of both requires the construction

of a 2x2 table for both CM and DP comparing the index test (HER2 IHC) against the reference standard (HER2 FISH). An example 2x2 table to determine the sensitivity and specificity of HER2 IHC is given below.

Table 6-2 Example 2x2 table

		FISH score (reference standard)	
		Amplified	Non-amplified
IHC score (index test)	+3	TP	FP
	+1	FN	TN

Better diagnostic tests will increase the number of true positive (TP) and true negative (TN) cases whilst simultaneously reducing false negative (FN) and false positive (FP) results that lead to incorrect treatment management. Sensitivity and specificity are calculated using the following equations:

$$\text{Sensitivity} = \text{TP}/(\text{TP} + \text{FN}) \quad \text{Specificity} = \text{TN}/(\text{TN} + \text{FP})$$

Values closer to 1 for each of these accuracy measures are ideal, demonstrating the ability of the index test to detect the true HER2 status. When sensitivity and specificity are calculated for each HER2 IHC method they can be compared to confirm the dominance of one over another.

Generally, the sensitivity and specificity of a particular test are estimated using data from individual studies, meta-analysis offers the opportunity to pool results from two or more studies into a single summary of effect (Haidich, 2010). Meta-analyses and systematic reviews are at the top of the pyramid of hierarchy of evidence (Murad et al., 2016), they provide high quality research evidence to inform decision-making in medicine. Meta-analyses of the sensitivity and specificity for each HER2 IHC method allows a more statistically robust comparison of accuracy than the use of single study estimates.

There are two challenges when summarising sensitivity and specificity into a single measure using meta-analyses approaches. First, the underlying trade-off between sensitivity and specificity (Irwig et al., 1995), as sensitivity increases, specificity decreases, thus they cannot be pooled without taking this into account. The second challenge is the use of scoring thresholds, these may vary across the individual studies from which accuracy results are extracted to be combined so that the HER2 status of a particular case may change across different thresholds.

Univariate meta-analysis approaches are not suitable when combining estimates of sensitivity and specificity since they pool these values separately ignoring the existing correlation between both (Takwoingi et al., 2017). Two approaches to meta-analysis exist that can be adopted for combining sensitivity and specificity estimates, these are the bivariate random effects model and the hierarchical summary receiver operating characteristic (HSROC) model (Menke, 2014) since both models allow for the correlation between sensitivity and specificity (Takwoingi et al., 2015). The bivariate model is used for meta-analysis if all included studies share a common underlying threshold (Takwoingi et al., 2015) otherwise the HSROC approach is appropriate as it allows the illustration of changes in sensitivity and specificity as the threshold varies (Rutter and Gatsonis, 2001).

Pathology labs are expected to follow guidelines when scoring HER2 IHC, these were set by ASCO/CAP and adopted in the UK. The accuracy measures determined for the IHC tests should reflect and match the most up to date thresholds to ensure they are relevant to current decision makers. Summary estimates of sensitivity and specificity can be reported for each of the 2008 and 2013 recommendations by subgrouping individual studies according to the thresholds that match the guidelines. The Cochrane handbook for systematic reviews of diagnostic test accuracy (Macaskill et al., 2010) recommends a bivariate model when studies are matched according to threshold.

6.1.7.2 Equivocal cases

Equivocal cases will not be included in the calculations of sensitivity and specificity since they are not represented in the 2x2 tables. Knowledge of their frequency is important when determining the superiority of one testing medium over another; they cause delays and add avoidable costs to the HER2 testing pathway. There are two groups of equivocal cases conditional on the true status, these must be separated in analyses as their proportions may differ. Once again, a meta-analysis summary estimate is superior to results taken from a single study but unlike sensitivity and specificity a univariate approach is sufficient. The probability of an equivocal score will be influenced by the underlying thresholds, so in the same way as for sensitivity and specificity they should first be sub-grouped by the threshold used prior to meta-analysis.

6.1.7.3 Current knowledge

Sensitivity, specificity and the probability of an equivocal score conditional on true status are the test performance measures of interest when comparing DP IA to CM for scoring HER2 IHC. A scoping search of the published literature was carried out to establish the extent of the available published evidence that reports the accuracy of HER2 IHC.

Dendukuri et al. (2007) and Cuadros and Villegas (2009) are two systematic reviews that summarised the level of agreement between the IHC and FISH tests for HER2. They both concluded that the IHC test was suitable for scoring negative HER2 cases but FISH was more reliable for the equivocal and positive cases. Unfortunately, both overlooked the thresholds and the approach taken to view the glass slides. A third review; a meta-analysis by Bahreini et al. (2015) pooled the concordance between IHC and FISH, proportion of false negatives, proportion of FISH positivity in IHC 2+ and 3+ score group. Again, they missed the effects of threshold and methods used. All three studies presented evidence of reduced accuracy with IHC compared to FISH but none addressed the issue of test performance with the different IHC testing methods.

A combined total of 58 studies were eligible for inclusion across the three reviews, out of these only two compared DP IA and CM for IHC scoring. Through citation and reference searching using these two studies, three further studies were identified comparing the two reporting mediums proving that studies exist from which the sensitivity and specificity for each IHC scoring medium can be determined.

However evidence of completed or ongoing reviews that attempt to pool the results from these and establish the accuracy of digitisation in comparison to the use of conventional microscopes when scoring HER2 IHC are not registered in PROSPERO, the Cochrane Library or the TRIP database. Unfortunately, there is not a robust evidence base in the literature to support the assertion of improved accuracy with DP IA in comparison to CM techniques.

6.2 Aims

This chapter aims to establish and compare the accuracy of digital pathology image analysis when compared to conventional microscopes by systematically reviewing the published literature and using meta-analysis methods to summarise accuracy data.

1. The primary aim is to generate and compare pooled estimates of accuracy when scoring HER2 IHC at the 2013 ASCO/CAP threshold levels
2. The secondary aims are to explore differences in the pooled accuracy estimates when 2008 thresholds are followed and to investigate differences in accuracy across the various digital technologies adopted

6.3 Methods

The Centre for Reviews and Disseminations guidance for carrying out reviews in healthcare (CRD, 2009) were followed to guide the systematic review process.

6.3.1 Information sources

A literature search was carried out to identify all the published head-to-head studies comparing HER2 IHC results using conventional and digital image analysis methods. Both bibliographic database searches and forensic searching techniques were carried out to identify all the relevant literature.

6.3.1.1 *Electronic searches*

The following databases were searched:

1. Medline
2. Embase
3. Web of Science
4. PubMed
5. Cochrane Library
6. Open Grey

Five key studies identified through the systematic reviews previously described, known to be relevant to the research aims were used to develop the search strategy. These were Dobson et al. (2010), Laurinaviciene et al. (2011), Minot et al. (2012), Mohammed et al. (2012) and Skaland et al. (2008). The key words were extracted from each paper and their titles and abstracts screened for any additional suitable search terms. These are listed in appendix D. Initially the search strategy combined five key concepts to describe invasive breast cancer patients tested for HER2, the digital pathology intervention as well as the conventional microscope comparator, the reference standard FISH and accuracy. Terms from the original key word list in appendix D were grouped under each of these key concepts. These formed the initial stage of the search strategy development and are given in appendix D.

Several iterations were made to the original strategy to arrive at the final version. The aim was to develop a strategy that retrieved the original five head-to-head studies listed above to be confident that it had identified all the available published evidence that met the study aims.

The search strategy was first developed in Medline and adapted for use in the remaining databases. The final search strategies for each of these are given in appendix F. These were used to conduct searches in each of the databases. A list of potential studies to be assessed

for inclusion in the systematic review was exported into EndNote and duplicates removed ready for screening.

6.3.1.2 Other sources

Forward and backward citation was carried out on all the eligible studies identified through the electronic searching process.

6.3.2 Eligibility criteria

6.3.2.1 Types of studies

All prospective and retrospective head-to-head studies comparing digital pathology image analysis to conventional microscope scoring of HER2 were eligible for inclusion in the review. All single arm studies were excluded to reduce the risk of bias due to confounding (Macaskill et al., 2010). Conference proceedings were as a rule excluded, however if they referred to a potentially eligible study, the corresponding full paper was retrieved where possible and screened against the inclusion criteria.

6.3.2.2 Types of participants

Studies that analysed the HER2 test results of women diagnosed with invasive breast cancer were eligible for inclusion in the review. The study population could either be randomly selected or comprise consecutive cases. This eliminated the risk of selection bias so that the participants were representative of the population of interest. It was noted in the scoping search prior to the electronic database searches that some studies over represent one or more of the HER2 scores or aim to include equal numbers of each in their study for analysis. Under these circumstances the study population did not represent the true prevalence of HER2 overexpression/amplification. These studies were excluded from the review. Studies that followed the testing pathways of patients diagnosed with cancers other than of the breast were also excluded.

6.3.2.3 Types of interventions

The intervention of interest was digital pathology with image analysis computer software to support the pathologists' reporting of HER2 IHC status. The study had to clearly specify the digital pathology platform and the image analysis software that were used to be eligible for inclusion in the review. Studies that use a digital pathology platform alone without image analysis software or do not mention the technology used were not eligible for inclusion in the review.

6.3.2.4 Types of comparators

Studies had to include a comparator arm representing standard practice in their analysis i.e. a pathologist reporting IHC HER2 status using a glass slide that is viewed under a microscope without the assistance of digital tools. Pathologists assign HER2 IHC status using pre-defined

scoring thresholds, to be eligible for inclusion in the review the study had to describe or give reference to the underlying scoring thresholds used by the pathologist. Where these were not given, the study was excluded from the review. Finally, the study had to provide the details of the assay used to stain the glass slide prior to HER2 reporting.

6.3.2.5 Type of reference standard

To be eligible, studies had to include a reference standard against which the HER2 IHC scores of the intervention and comparator arms can be compared. The reference standard must be FISH since it is considered the gold standard for HER2 scoring. Studies that did not include a reference standard or used other methods such as consensus scoring to confirm HER2 status are excluded from the review. In addition, the study had to specify the FISH assay used and the underlying scoring threshold followed by the pathologist.

In summary, all eligible studies include a sample of glass slides with tissue specimens taken from a group of women diagnosed with invasive breast cancer. Each slide is subjected to three versions of HER2 testing, IHC digital, IHC manual and FISH as the reference standard so that each invasive breast cancer patient has a score for each method.

6.3.2.6 Types of accuracy measures

There were four accuracy measures of interest; sensitivity, specificity, the probability of an equivocal score if true status is negative and the probability of an equivocal score if true status is positive. The reference standard HER2 results had to be compared against both the intervention and the comparator results independently. Studies were eligible for inclusion in the review if they presented the raw data needed to populate a 2x2 table for each arm of the study plus the number of equivocal cases for each IHC method split between those that are FISH positive and FISH negative. Where a study met all the eligibility criteria but there were gaps in the data described, study authors were contacted. Where access to missing data was granted the study was considered eligible, otherwise it was excluded from the review.

6.3.2.7 Study limits

A limit on the publication date was not set as thresholds for scoring HER2 have altered in the past 10 years. Rather all studies comparing DP IA and CM to FISH were included but grouped by the threshold used for scoring for data analysis. Only English language publications were included.

Table 6-3 Systematic review of accuracy - summary inclusion criteria

	Include	Exclude
Population	Women diagnosed with invasive breast cancer that are tested for HER2	Conditions other than breast cancer
Intervention	HER2 IHC digital pathology image analysis	No Image Analysis
Comparator	HER2 IHC conventional microscope	None or other comparator
Reference standard	HER2 FISH	None or other reference standard
Accuracy measure	Raw data available to populate 2x2 tables for intervention and comparator	Data missing and unobtainable
Study Design	Retrospective and prospective studies	-----
Language	English only publications	Non-English publications
Publication Date	None	None

6.3.3 Data collection and analysis

6.3.3.1 Study selection

Study selection followed a two-step process and involved two reviewers. To start; the first reviewer screened all titles and abstracts against the pre-specified inclusion criteria. Where it was obvious that papers did not meet the review objectives they were excluded at this stage. The full papers of the remaining studies were retrieved for a second round of screening. These were screened by the first reviewer against the inclusion criteria, publications that were not eligible were excluded. The final list of studies were considered eligible for inclusion in the systematic review.

A second reviewer screened all papers that were included by the first reviewer as well as screening a random subset of excluded papers to ascertain that the first reviewer made no inappropriate exclusions. Disagreement between the two reviewers were mediated through consultation with a third reviewer.

6.3.3.2 Data extraction

Data was extracted from the eligible studies using a pre-designed data extraction form given in appendix G. This included information on the thresholds used for subjective scoring of conventional IHC and quantitative FISH scoring including the assays used for staining slides. For digital systems; information on the scanner and methods of image analysis were extracted. Finally, the data needed to populate the 2x2 tables and to calculate the probability

of an equivocal score for each IHC method were extracted from each included study. Where the study generated results to populate more than one 2x2 table, each was treated as an independent study.

6.3.3.3 Reporting quality

All eligible studies were quality assessed using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool (Whiting et al., 2011). The signalling question related to the interval between the index test and reference standard was not relevant so was excluded. The questions related to how the cases were selected, blinding of results and thresholds described were all included. The findings from each study were summarised and compared to identify any trends in the risk of bias. All primary studies were included in the meta-analysis regardless of risk of bias or applicability concerns. Where appropriate in the sensitivity analysis, studies are grouped according to their QUADAS-2 outcomes to identify any association between these and the accuracy result reported.

6.3.4 Data synthesis

For each study the scoring thresholds used for the IHC test on conventional microscope and for the FISH test as well as the technology and software used for the DP IA arm were recorded. The studies were organised according to the thresholds used into three groups:

1. Comply with 2008 ASCO/CAP guidelines
2. Comply with 2013 ASCO/CAP guidelines
3. A miscellaneous group that do not match either of the two sets of recommendations.

It was anticipated that across the eligible studies the technology used for the DP IA arm will differ. Developers of the technology have incorporated their own unique underlying methods for scoring HER2 positivity which they convert to the conventional 0, +1, +2 and +3 scores. However many of these are still in development and the systematic review was not expected to retrieve many studies that matched on both thresholds used and technology to take account of this second layer of sub-division. For the main analyses, studies were grouped by thresholds only, if enough studies were retrieved that match on both threshold and technology, a sensitivity analysis scenario will explore accuracy outcomes using this grouping.

6.3.4.1 Base case analysis

The base case analysis (BCA) represented current HER2 scoring recommendations and included the studies that follow the 2013 ASCO/CAP guidelines.

6.3.4.2 Sensitivity analysis

Sensitivity analysis (SA) explored changes in estimates of accuracy with different scoring threshold and technology adopted in the digital arm. In summary, the accuracy under three scenarios was investigated:

1. SA1: Studies that adhere to the 2008 ASCO/CAP guidelines for scoring were grouped and data from these pooled to generate accuracy estimates at the 2008 threshold
2. SA2: If appropriate the studies used in the base case analysis were further sub-divided according to the digital technology used and analysed independently.
3. SA3: The final scenario sub-divides all eligible studies by DP IA technology and groups them accordingly. A bivariate meta-analysis will explore differences in sensitivity and specificity estimates with each technology used

6.3.4.3 Meta-analysis

Across the base case and sensitivity analysis scenarios four indicators of test performance with 95% confidence intervals (CI) were estimated for each of CM and DP IA for IHC testing:

1. Sensitivity
2. Specificity
3. Probability of equivocal status conditional on true positive HER2
4. Probability of equivocal status conditional on true negative HER2

6.3.4.3.1 Sensitivity and specificity

Sensitivity and specificity were calculated for each of the eligible studies using RevMan software and presented in a forest plot for illustrative purposes only, they were not pooled at this stage.

Raw data from each individual study was entered into R software and the *Mada* (Doebler and Holling, 2015) package used to carry out the bivariate meta-analysis of sensitivity and specificity. Point estimates with 95% confidence intervals were generated and summary ROC curves were used to compare accuracy outputs across the two IHC testing methods.

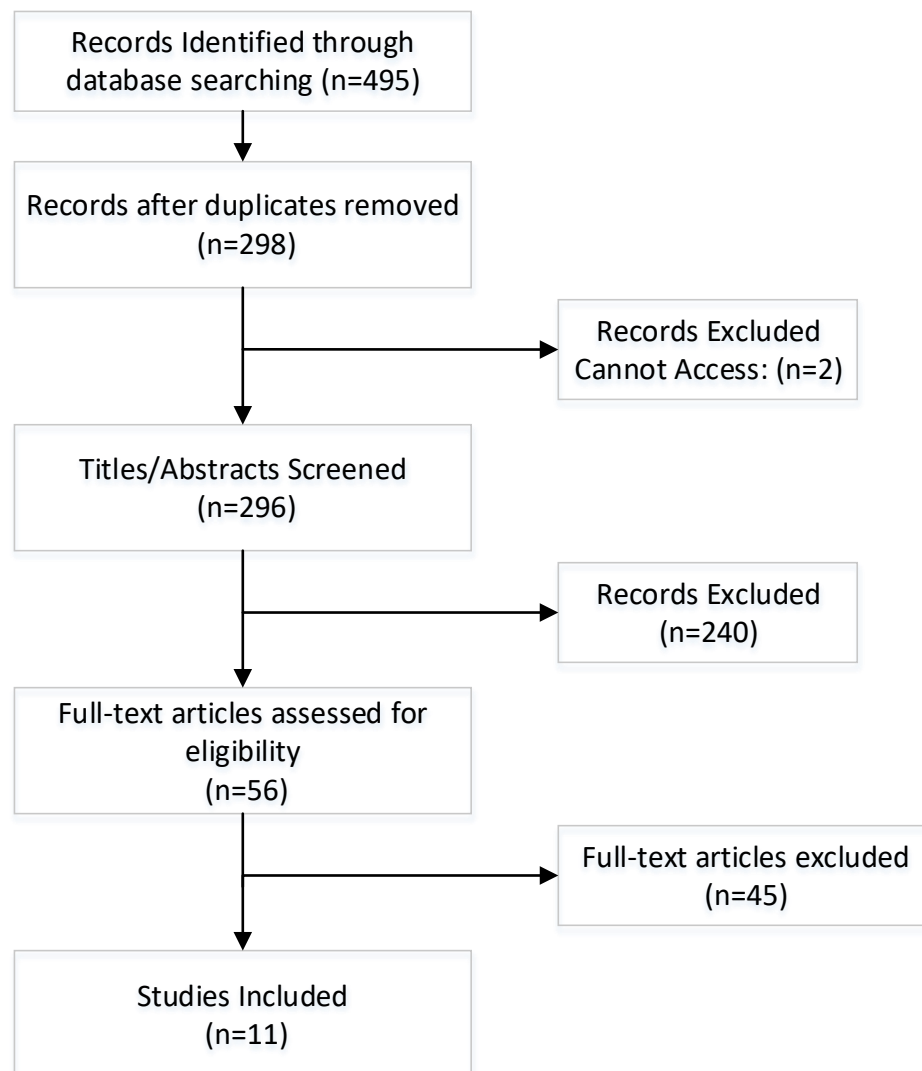
6.3.4.3.2 Equivocal scores

A random effects univariate meta-analysis was carried out to pool the proportion of equivocal scores conditional on true status for each IHC testing method. This was completed using the *Meta* (Schwarzer, 2007) package in R for all of the base case and the first and second sensitivity analysis scenarios.

6.4 Results

A total of 495 publications were identified through the electronic database searches, 240 were excluded when reviewing the titles and abstracts. Access was not available for two of the studies, an abstract was available for Bishop et al. (2002) but not for Judd et al. (2001), based on this information alone it was not possible to make an eligibility decision so they were both excluded from the review. 45 studies were excluded on the second round of screening the full papers. The Preferred Reporting Items for Systematic Review and Meta-analysis (PRISMA) flowchart (Moher et al., 2015) summarising the process for the identification and selection of studies is given in figure 6-3.

Figure 6-3 Study identification and selection process



The 45 full text articles were excluded for a number of reasons, including:

- 8 studies deliberately over representing one of the IHC HER2 outcomes or included an equal number for each category.

- 8 papers had issues with their reference standard;
 - 5 did not carry out FISH testing on all cases and
 - 3 directly compared CM and DP IA without including a reference standard
- 15 studies had an issue with the index test or comparator or both;
 - 6 compared DP IA to FISH with no CM,
 - 7 compared CM to FISH with no DP IA and
 - 2 papers didn't carry out DP IA by IHC but compared DP IA by FISH to standard FISH.
- 4 studies did not include enough data to populate 2x2 tables and so it was not possible to calculate the sensitivity and specificity for these
- 1 paper was a non-English publication
- 1 paper did not specify the thresholds used to score both IHC and FISH
- 1 paper was a narrative review and so did not present any relevant data
- 7 records were conference abstracts which were included initially to establish if they presented any new information that could be included in the meta-analysis.
 - 2 of the conference abstracts were related to two published papers already included in the review and
 - The 5 remaining did not include enough data in the abstract to populate 2x2 tables. Published papers related to these conference abstracts were sought but without success. All five were excluded.

It is worth highlighting that all known five head-to-head studies were retrieved by the search strategy and met the inclusion criteria. These made up nearly 50% of the studies included in the systematic review. An overview of the reference list of all included papers and a citation search identified no further relevant evidence to be included in the review.

6.4.1 Study characteristics

Table 6-4 gives an overview of the assays and scoring thresholds used in the included studies. Eight out of the eleven studies used Herceptest (DAKO) for IHC staining on all or a group of their slides. Dobson et al. (2010) sample was made up of consecutive cases from two hospitals where one used both Herceptest (DAKO) and Oracle (Leica) and the second used Pathway (Ventana) for HER2 staining. Also Joshi et al. (2007) used both Herceptest (DAKO) and Pathway (Ventana). Finally Wang et al. (2001) uses a different antibody but from the same manufacturers as the HercepTest. The IHC assay used is the same across both CM and DP IA since the same stained slides are viewed under a microscope and scanned and analysed

digitally. A similar pattern can be seen across the studies for the FISH analysis, 10 out of the 11 studies used PathVysion (Abbott) as their preferred assay.

When considering the accuracy of the CM scoring method to report on HER2 status there are two levels of threshold variability; one at the subjective IHC scoring level and the second when confirming FISH status. For each study these are compared to the ASCO/CAP scoring recommendations of 2008 and 2013 as described by Nitta et al. (2016). The final column in table 6-4 summarises this; giving the year of the ASCO/CAP guidelines to which each paper's scoring threshold for subjective IHC and FISH match. Only 6 out of the 11 studies adhere to either the 2008 or 2013 thresholds when scoring, the remaining studies use alternative combinations of FISH and IHC thresholds to define positivity.

For the DP IA arm in each review there are two steps to consider; firstly the method and equipment used for digitisation and secondly the algorithm used to analyse the digital slides and give a HER2 score. These are summarised in table 6-5. ACIS is the most commonly used platform for both digitisation and analysis; this system was used by 4 of the included studies. Dobson et al. (2010) and Mohammed et al. (2012) both use SlidePath for digitisation but the former use their own IA algorithm and the latter using the SlidePath system. Two studies use the Ariol system, one uses MDS and one uses Visiopharm technology for both digitisation and analysis. Finally Joshi et al. (2007) use either a scanner or microscope for generating a digital image which is analysed using the authors developed algorithms.

Table 6-4 Summary of included studies

Study	Sample Size	Country of Research	IHC		FISH		ASCO/CAP Guidelines ¹
			Assay	Positivity Threshold	Assay	Positivity Threshold	
Bloom and Harrington (2004)	129	USA	HercepTest	>10%	Inform & PathVysion	≥2.0	2013
Ciampa et al. (2006)	108	USA	HercepTest	>10%	PathVysion	≥2.0	2013
Dobson et al. (2010)	136	Ireland	HercepTest & Oracle & Pathway	>30%	PathVysion	>2.2	2008
Ellis et al. (2005)	114	UK	HercepTest	>10%	PathVysion	>2.2	NA
Joshi et al. (2007)	450	USA	HercepTest & Pathway	>10%	PathVysion	≥2.2	NA
Laurinaviciene et al. (2011)	152	Lithuania	Pathway	>10%	PathVysion	>2	2013
Minot et al. (2012)	154	USA	HercepTest	>30%	PathVysion	>2.2	2008
Mohammed et al. (2012)	431	UK	HercepTest	>10%	Vysis	>2	2013
Skaland et al. (2008)	219	Norway	HercepTest	>10%	PathVysion	>2.2	NA
Turashvili et al. (2009)	579	USA	Pathway	>10%	PathVysion	>2.2	NA
Wang et al. (2001)	189	USA	A0485	>50%	PathVysion	≥ 2	NA

Table 6-5 Summary digital pathology image analysis methods adopted in each study

Study	Digitisation method	Image Analysis Algorithm
Bloom and Harrington (2004)	ACIS digital microscope	An ACIS assisted HER2 score is generated by the system software
Ciampa et al. (2006)	The ACIS system includes an automated robotic bright field microscope and a windows NT-based computer. The microscope scans the slide and the computer generates a digital image	ACIS system generates a HER2 score
Dobson et al. (2010)	SlidePath using a NanoZoomer DP system to digitize the slides	IA algorithm developed by authors and used within SlidePath's tissue IA system
Ellis et al. (2005)	Micrometastasis detection system (MDS) is a computer controlled scanning microscope	A digital scoring application developed for MDS
Joshi et al. (2007)	Either a scanner or microscope with a digital camera attached	Algorithm developed by authors
Laurinaviciene et al. (2011)	ArrayImager software module from Visopharm	HER2-CONNECT software module from Visopharm
Minot et al. (2012)	ACIS	ACIS assisted assessment
Mohammed et al. (2012)	Slides are scanned with Hamamatsu NanoZoomer	SlidePath Tissue Image analysis system
Skaland et al. (2008)	Ariol SL 50 system made up of an automatic slide loader and scanner	Paper states IA algorithm applied.

		Does not indicate if this was part of the DP system or developed by authors
Turashvili et al. (2009)	Ariol system based on an Olympus microscope and equipped with a black and white video camera	Ariol IA software automatically generates a score
Wang et al. (2001)	ACIS consisted of an automated robotic bright-field microscope module, a computer, and a Windows NT-based software interface. Robotic microscope scanned slides. Computer monitor displayed digital images	ACIS able to quantitate HER2 staining to determine HER2 status

6.4.2 Quality assessment














































































The quality of the studies included in the review were evaluated for bias and applicability concerns using the QUADAS-2 tool. A summary of the results is given in table 6-6. All except one study performed well on applicability confirming that the primary studies included are applicable to the research question (Whiting et al., 2011). Dobson et al. (2010) is the only study where there was some concern around the applicability of patient selection but this was due to lack of information rather than a clear problem. Even though the study mentions invasive breast cancer extensively in the introduction this is not explicitly stated in the methods or results section when describing the study sample. Dobson et al. (2010) simply refer to cases without describing the types of cases included. Other than this minor issue there were no other concerns regarding applicability across all published studies included in the review.

The risk of bias was variable both across studies and across the criteria. Bias due to both the performance of the reference standard and index test was of the highest concern followed by patient selection and flow and timing. 7 out of the 11 studies scored poorly on both the index test and reference standard components. The scoring thresholds were clearly stated and the same reference standard i.e. FISH testing was used in all the studies as these are both fundamental criteria for inclusion in the review. The main concern in both arms was failure on blinding. Either it was obvious that the tests were interpreted with prior knowledge

of the other test results or it was not clear if blinding was considered at all. Blinding of test results has been reported elsewhere as a design feature of diagnostic accuracy studies that consistently introduces bias (Furukawa and Guyatt, 2006), the studies included support this finding.

Under patient selection, there was some concern around the sample selection process. It was unclear in 4 of the studies how this was carried out raising concerns about potential selection bias. This occurs when eligible cases are not selected consecutively or at random (Roever, 2015). Finally, out of the 4 domains to assess risk of bias; published studies in most cases scored positively on flow and timing. Dobson et al. (2010) and Turashvili et al. (2009) had issues around data exclusion; it was not clear or was obvious that data was excluded but no justification for this was given.

Table 6-6 Quality assessment of included studies

Study	ASCO/CAP guidelines	RISK OF BIAS				APPLICABILITY CONCERNS		
		PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD
Bloom and Harrington (2004)	2013							
Ciampa et al. (2006)	2013							
Dobson et al. (2010)	2008							
Ellis et al. (2005)	None							
Joshi et al. (2007)	None							
Laurinaviciene et al. (2011)	2013							
Minot et al. (2012)	2008							
Mohammed et al. (2012)	2013							
Skaland et al. (2008)	None							
Turashvili et al. (2009)	None							
Wang et al. (2001)	None							

6.4.3 Accuracy data extraction

The main aim of the review was to summarise estimates of sensitivity and specificity for DP IA and compare these against those for CM. Data to populate 2x2 tables for each arm of the analysis were extracted from the eligible studies. The individual studies presented enough data to populate at least one pair of 2x2 tables. Data was available to populate more than one 2x2 table in the following studies:

1. Bloom and Harrington (2004) had ten independent pathologists score IHC slides using both a microscope and the ACIS digital pathology system. This study generated 10 pairs of 2x2 tables to be included in the review.
2. Joshi et al. (2007) included two groups of breast cancer cases; those set up as WMDs (WS) and a second group of cases as Tissue Microarrays (TMA). Data to populate two 2x2 tables representing each of the WS and TMA groups were available for the CM arm of the analysis. For the DP IA arm they presented two methods of scoring; an ordinal algorithm (OA) and a continuous algorithm (CA). These were applied to both the WS and TMAs giving data to populate four 2x2 tables to represent DP IA accuracy.
3. For Laurinaviciene et al. (2011) the same pathologist performed the subjective microscope evaluation twice but with a two month interval. For this paper two 2x2 tables were generated for CM and one for DP IA.
4. Minot et al. (2012) had 3 pathologists score manually and 3 cytotechnologist (CyT) score both manually and digitally. The Pathologists scores were excluded since there was no digital equivalent included in the review. This paper generated three pairs of 2x2 tables, one for each CyT to be included in the review.
5. Skaland et al. (2008) subjective IHC scoring was carried out by 3 independent pathologists giving 3 sets of 2x2 tables for CM. They also presented a subjective consensus score but this was excluded from the analysis since in reality it is not common practice for more than one clinician to be available when scoring HER2. Only one 2x2 table is generated for DP IA.
6. Turashvili et al. (2009) study has two independent pathologist carry out subjective scoring using CM. This study gives two 2x2 tables for the CM arm of the analysis and one for the DP IA arm.

In total there are 11 publications to be included in the review giving enough data to populate 27 2x2 tables for CM and 25 2x2 tables for DP IA. The 2x2 tables are used to evaluate the concordance between IHC and FISH and so determine the difference in accuracy between conventional subjective methods and quantitative digital techniques. From each 2x2 table

the sensitivity and specificity were calculated for each study. For the same number of 2x2 tables it was also possible to calculate the split between cases scored as equivocal and those giving a definite score with IHC.

Improvements in the accuracy of IHC HER2 scoring are characterised by increases in both sensitivity and specificity together with a reduction in the number of equivocal scores generated. Table 6-7 summarises the direction of change when comparing accuracy point estimates for DP IA against those for CM. More than one set of outputs are generated for many of the studies as described on page 160.

There does not appear to be a general trend to support the superiority of digital methods over conventional techniques. In 14 out of the 27 comparisons, there is a reduction in the sensitivity of IHC scoring with digital and in 11 cases there is also a reduction in specificity. An improvement in the proportion of 2+ scores is observed in 16 of the summary point pairs given.

Forest plots of the sensitivity and specificity for CM and DP IA scoring are given in figure 6-4 and 6-5 respectively. There appears to be a greater level of uncertainty around sensitivity in both the CM and DP IA data due to the wider confidence intervals, in addition there is also evidence of greater variability in the sensitivity than in specificity estimates due to the scatter of the points. The variability in specificity estimations decreases with digital pathology but it is not so clear if there is a difference in sensitivity between microscope and digital techniques.

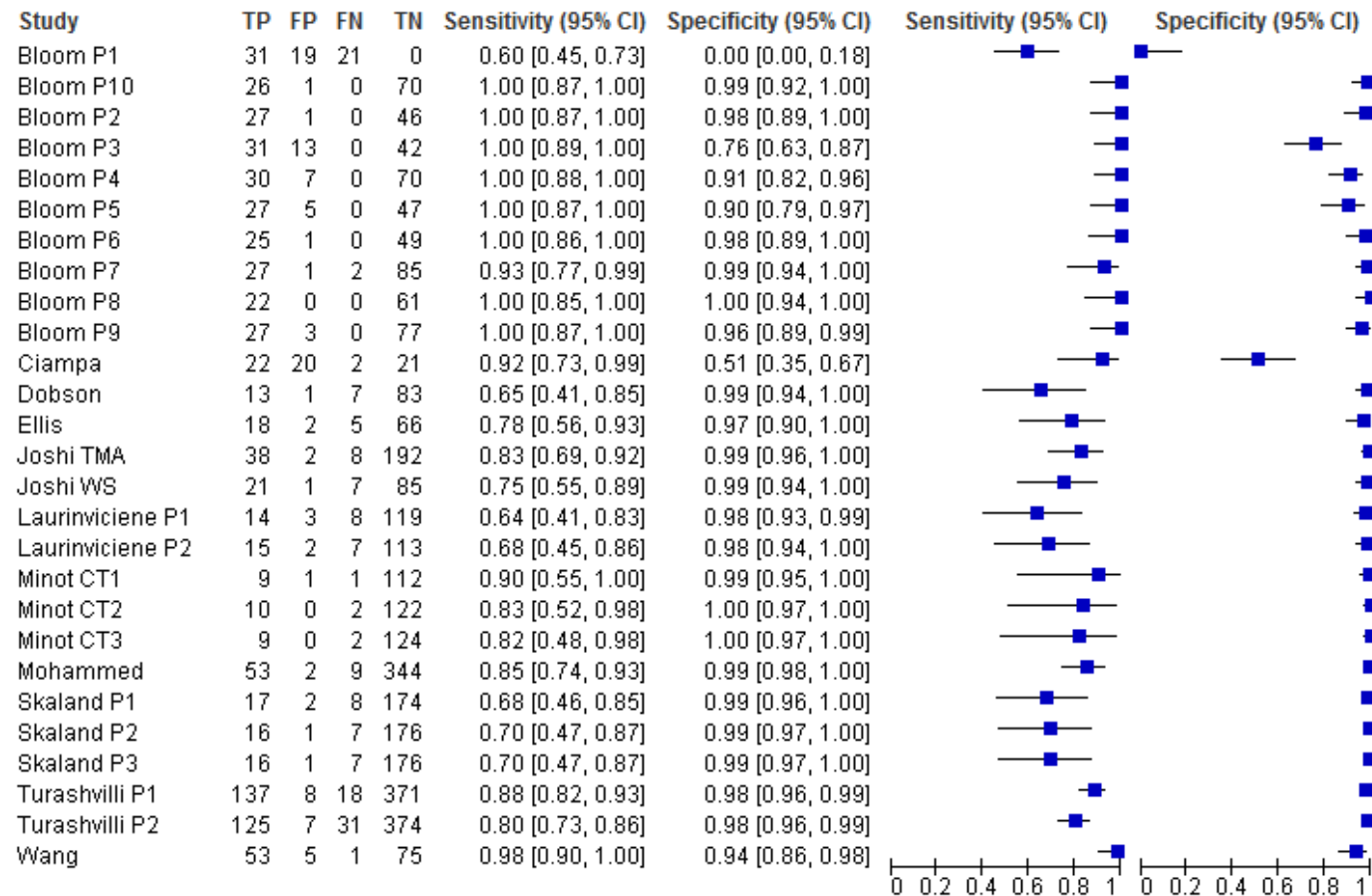
Based on this data alone there is not enough evidence to conclude the ability of DP IA to improve the accuracy of IHC reporting and so meta-analyses were carried out.

Table 6-7 Summary change in point estimates of sensitivity (SN), specificity (SP) and proportion of equivocal results reported in each study when comparing DP to CM

Studies		Accuracy Δ		
		SN	SP	%2+
Bloom and Harrington (2004)	P*1	↓	↑	↓
	P2	↓	↓	↓
	P3	↓	↑	↓
	P4	↓	↑	↓
	P5	↓	↑	↓
	P6	↓	↓	↓
	P7	↑	↓	↓
	P8	↓	↓	↓
	P9	↓	↓	↓
	P10	↓	↓	↓
Ciampa et al. (2006)		↑	↑	↑
Dobson et al. (2010)		↑	↓	↓
Ellis et al. (2005)		↑	----	↓
Joshi et al. (2007)	WSOA	↑	----	↑
	WSCT	↑	↑	↑
	TMA OA	↑	----	↑
	TMA CT	↑	↑	↑
Laurinaviciene et al. (2011)	P1	↑	↓	↑
	P2	↑	↓	↑
Minot et al. (2012)	CyT1	↓	----	↓
	CyT2	----	----	↓
	CyT3	↑	↓	----
Mohammed et al. (2012)		↓	----	↓
Skaland et al. (2008)	P1	↑	----	↑
	P2	↑	----	↑
	P3	↑	----	↑
Turashvili et al. (2009)	P1	↓	↑	↑
	P2	↓	↑	↑
Wang et al. (2001)		↓	↓	↓

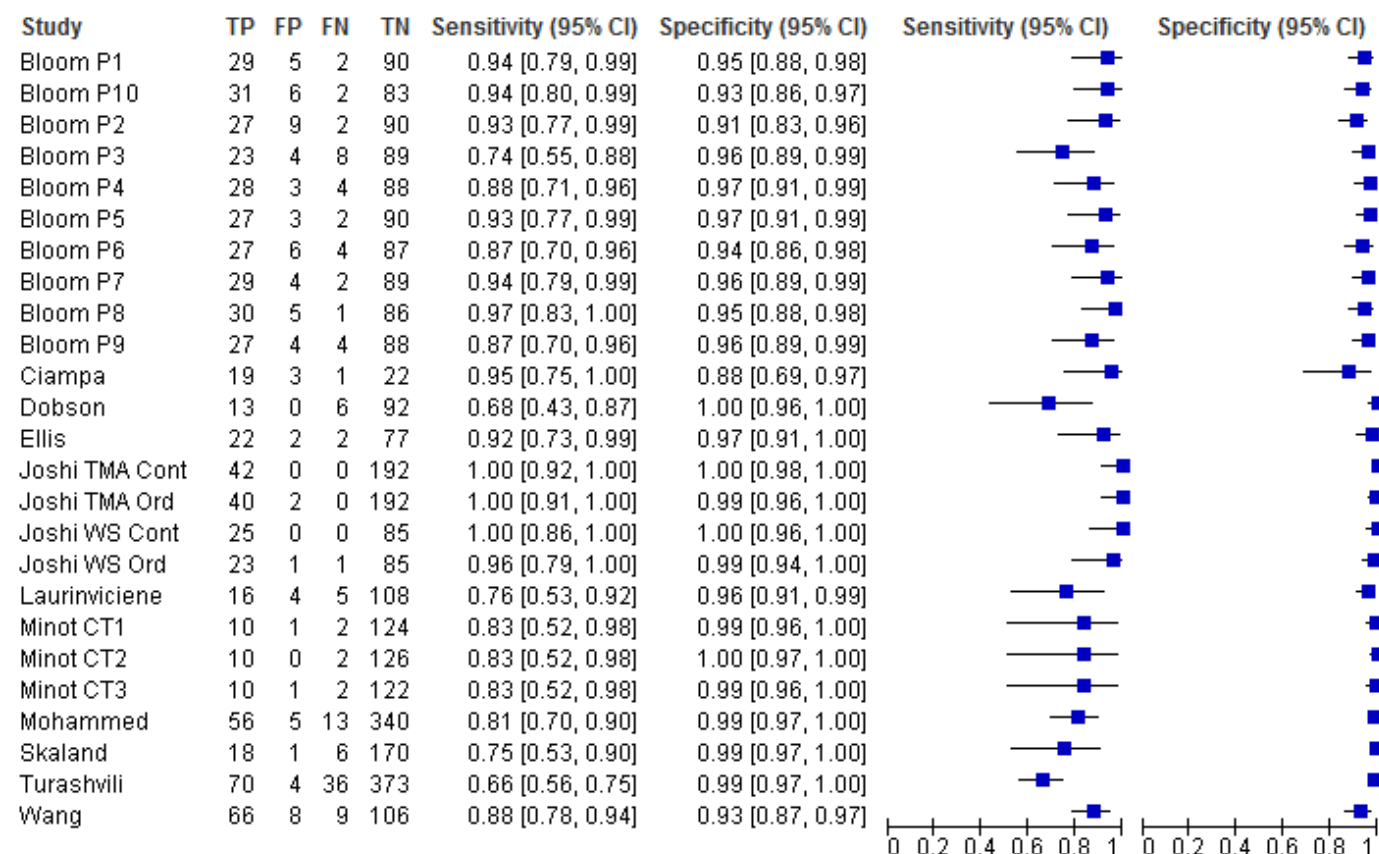
↑ / ↓ Favourable for DP ↓ / ↑ Favourable for CM ---- No Change (P refers to the pathologist)

Figure 6-4 Sensitivity and specificity of HER2 IHC using CM



Studies are included more than once in the forest plot where data was available to populate more than one 2x2 table as described on pg162
P-Pathologist TMA-Tissue Microarrays WS-Whole Slides CyT-Cytotechnologist

Figure 6-5 Sensitivity and specificity of HER2 IHC using DP IA



Studies are included more than once in the forest plot where data was available to populate more than one 2x2 table as described on pg162
P-Pathologist TMA-Tissue Microarrays WS-Whole Slides Ord-Ordinal Algorithm Cont-Continuous Algorithm CyT-Cytotechnologist

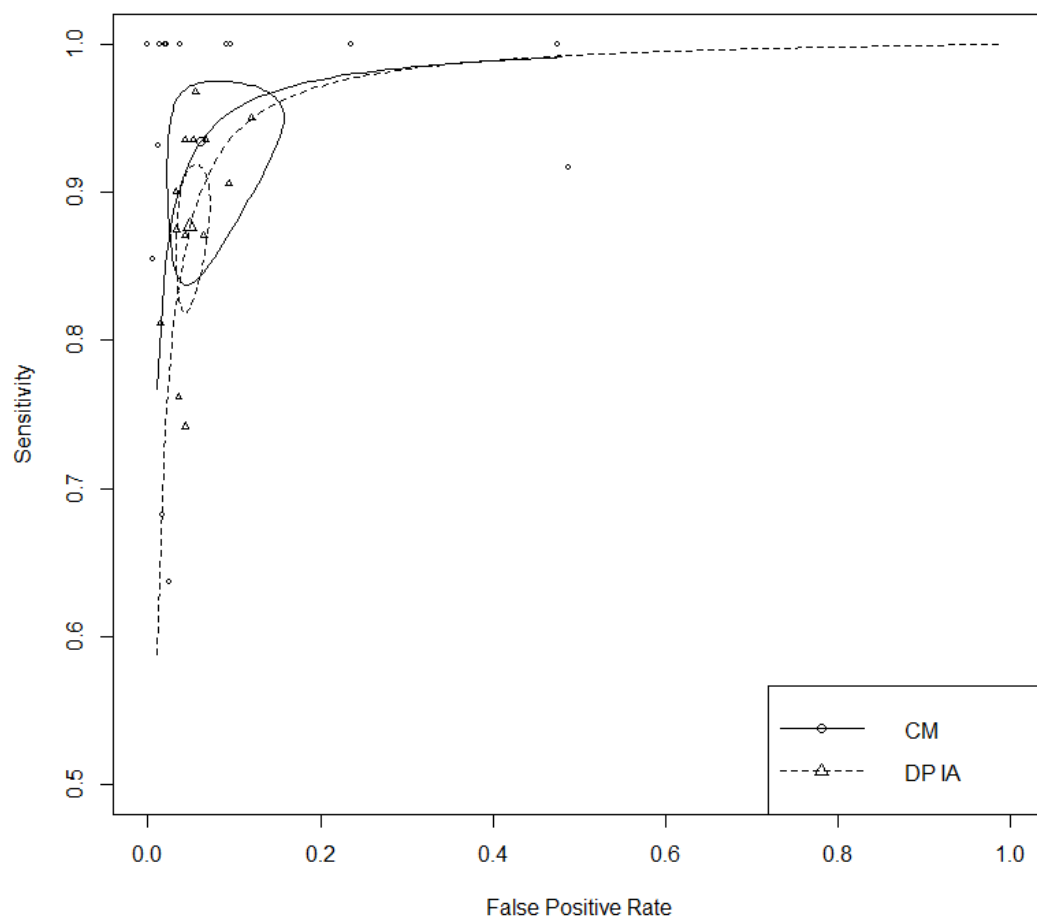
6.4.4 Base case analysis

Four studies met the criteria of the ASCO/CAP 2013 guidelines; data extracted from these were included in the base case meta-analyses. These papers were:

1. Bloom and Harrington (2004)
2. Ciampa et al. (2006)
3. Laurinaviciene et al. (2011)
4. Mohammed et al. (2012)

The bivariate meta-analysis for base case analysis estimated sensitivity as 0.934 (95%CI: 0.863, 0.970) and the false positive rate (FPR) as 0.061 (95%CI: 0.027, 0.131) for the CM arm. For DP IA, sensitivity was estimated 0.877 (95%CI: 0.832, 0.912) and the FPR as 0.049 (95%CI: 0.036, 0.067). The point estimates suggests a decrease in sensitivity with DP IA with a simultaneous decrease in the FPR denoting an increase in specificity. The summary Roc curve for the base case analysis comparing both IHC scoring methods is given in figure 6-6.

Figure 6-6 BCA comparison of CM and DP IA with SROC curves



The confidence regions for sensitivity and specificity overlap but cover a tighter area for the DP IA data suggesting a lower level of uncertainty around the predictions with digitisation.

Meta-analysis of the proportion of equivocal results conditional on true status were also carried out as part of the base case analysis. Forest plots for the HER2 positive cases across each IHC method are given in figures 6-7 and 6-8 and the HER2 negative cases in figures 6-9 and 6-10.

Figure 6-7 BCA proportions of equivocal scores if HER2 positive: CM

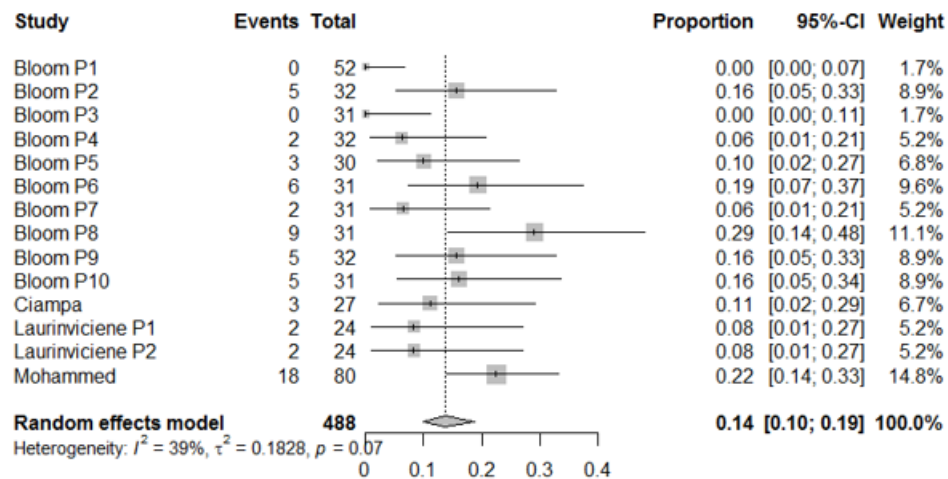
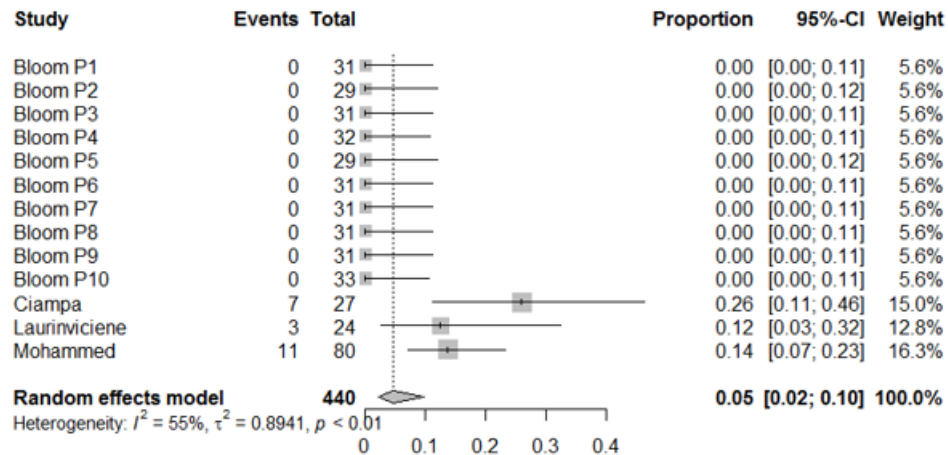


Figure 6-8 BCA proportion of equivocal scores if HER2 positive: DP IA



Across both the positive and negative arms there appears to be a statistically significant reduction in the proportion of HER2 IHC tests scored as equivocal when going from the CM to DP IA approach. However high levels of heterogeneity exist especially for the comparison across the HER2 negative cases due to high value of I^2 accompanied with a low p-value.

Figure 6-9 BCA Proportion of equivocal scores if HER2 negative: CM

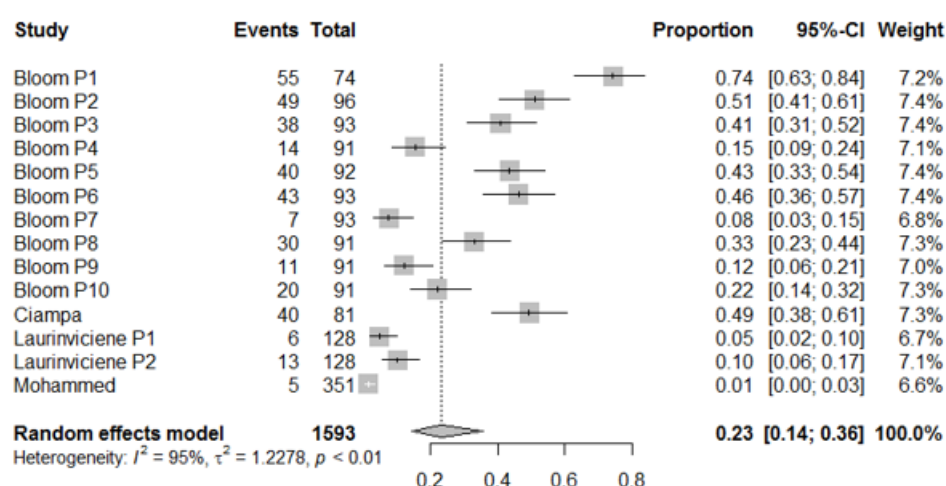
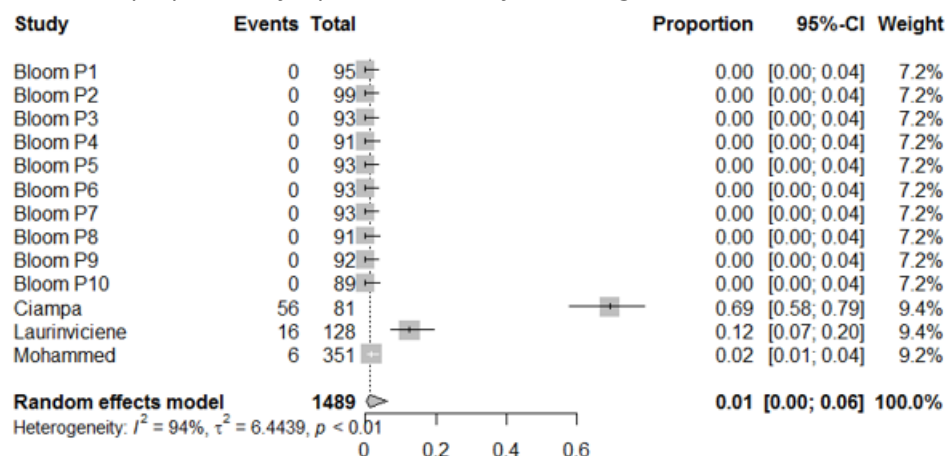


Figure 6-10 BCA proportion of equivocal scores if HER2 negative: DP IA



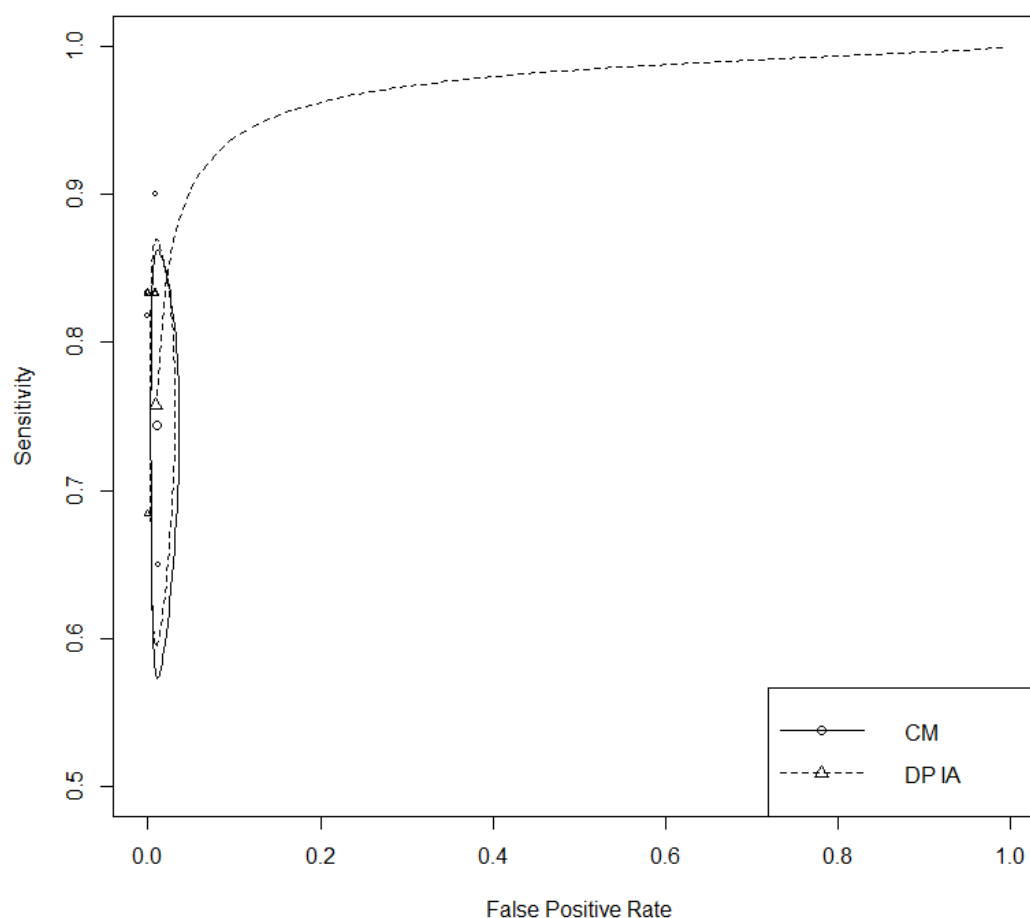
6.4.5 Sensitivity analysis

6.4.5.1 SA1 2008 thresholds

The base case analysis limited the studies to only include those that met the 2013 guidelines. The first sensitivity analysis explored the impacts on the accuracy measures when analysing data from studies that base their scoring on the 2008 guidelines. There were 2 studies that met this criteria, Dobson et al. (2010) and Minot et al. (2012).

Under this scenario for CM the sensitivity was determined to be 0.744 (95%CI: 0.610, 0.843) and the false positive rate as 0.011 (95%CI: 0.004, 0.029). For DP IA the estimate of sensitivity was 0.758 (95%CI: 0.630, 0.851) and for the FPR 0.009 (95%CI: 0.004, 0.025). The point estimates suggest that sensitivity has increased and so has the specificity owing to an increase in the FPR, however once again there is considerable overlap between the confidence intervals. The SROC in figure 6-11 supports this notion since both confidence regions almost overlap each other, signifying no difference in accuracy between each HER2 IHC scoring method.

Figure 6-11 SA1 Comparison of CM and DP IA with SROC curves



The differences in the proportion of equivocal scores was also meta-analysed under this scenario, the results for the HER2 positive cases are given in figures 6-12 and 6-13 and for the negative cases in figures 6-14 and 6-15. There is evidence of a reduction in equivocal scores for both HER2 positive and negative cases but these may not be statistically significant due to the overlap of confidence intervals. Unlike the BCA, lower levels of heterogeneity exist when pooling the studies but in the same way as for the BCA, heterogeneity is lower across analyses for positive HER2 cases.

Figure 6-12 SA1 Proportion of equivocal cases if HER2 positive: CM

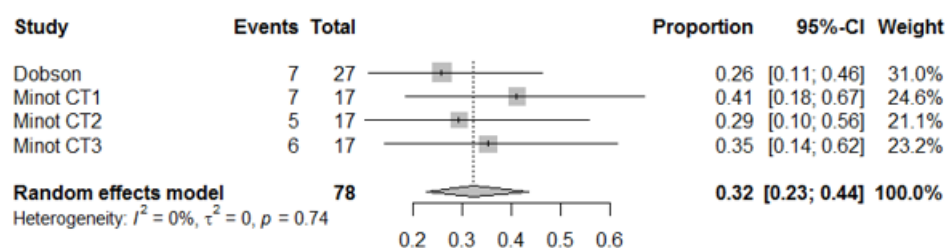


Figure 6-13 SA1 Proportion of equivocal cases if HER2 positive: DP IA

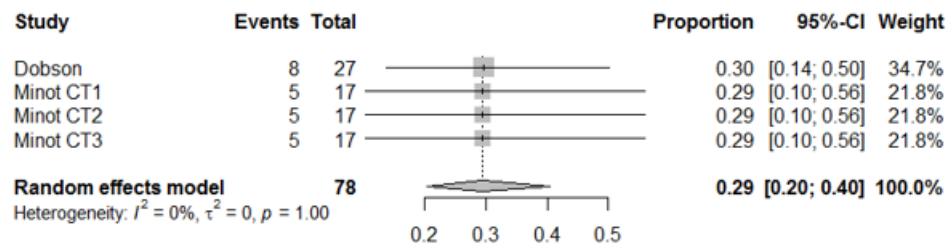


Figure 6-14 SA1 Proportion of equivocal cases if HER2 negative: CM

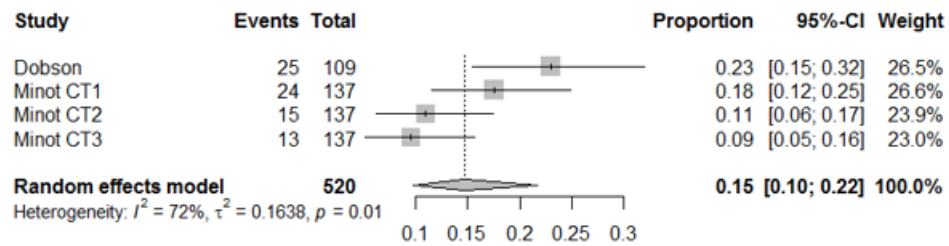
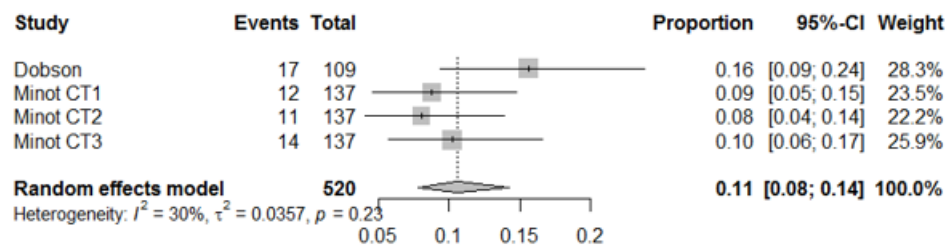


Figure 6-15 SA1 Proportion of equivocal cases if HER2 negative: DP IA



6.4.5.2 SA2 Thresholds and DP IA technology matched

The scanning and scoring technology varied across the studies included in the BCA; two studies used ACIS technology, one VisioPharm and the last SlidePath. This scenario summarised test performance using data from the two studies using ACIS technology; Bloom and Harrington (2004) and Ciampa et al. (2006). The studies matched on thresholds used for scoring the CM IHC test and the DP IA technology used.

Under this scenario for the CM analysis, sensitivity was 0.960 (95%CI: 0.925, 0.979) and the FPR was 0.086 (95%CI: 0.035, 0.192). For DP IA the point estimate for sensitivity decreased to 0.900 (95%CI: 0.850, 0.935) and the point estimate for the FPR also decreased to 0.059 (95%CI: 0.045, 0.079) signifying an increase in specificity. The SROC curve in figure 6-16 suggests sensitivity decreases with DP IA even though there is some overlap between the two confidence regions and for specificity, DP IA appears to more precisely identify negative HER2 cases hence the narrower confidence region for the FPR i.e. specificity. Similar trends for equivocal scores as those seen under BCA are observed under this scenario. There is a

statistically significant reduction in the proportion of HER2 equivocal IHC cases however these estimates are coupled with high levels of heterogeneity due to high value of I^2 accompanied with a low p-value.

Figure 6-16 SA2 Comparison of CM and DP IA with SROC curves

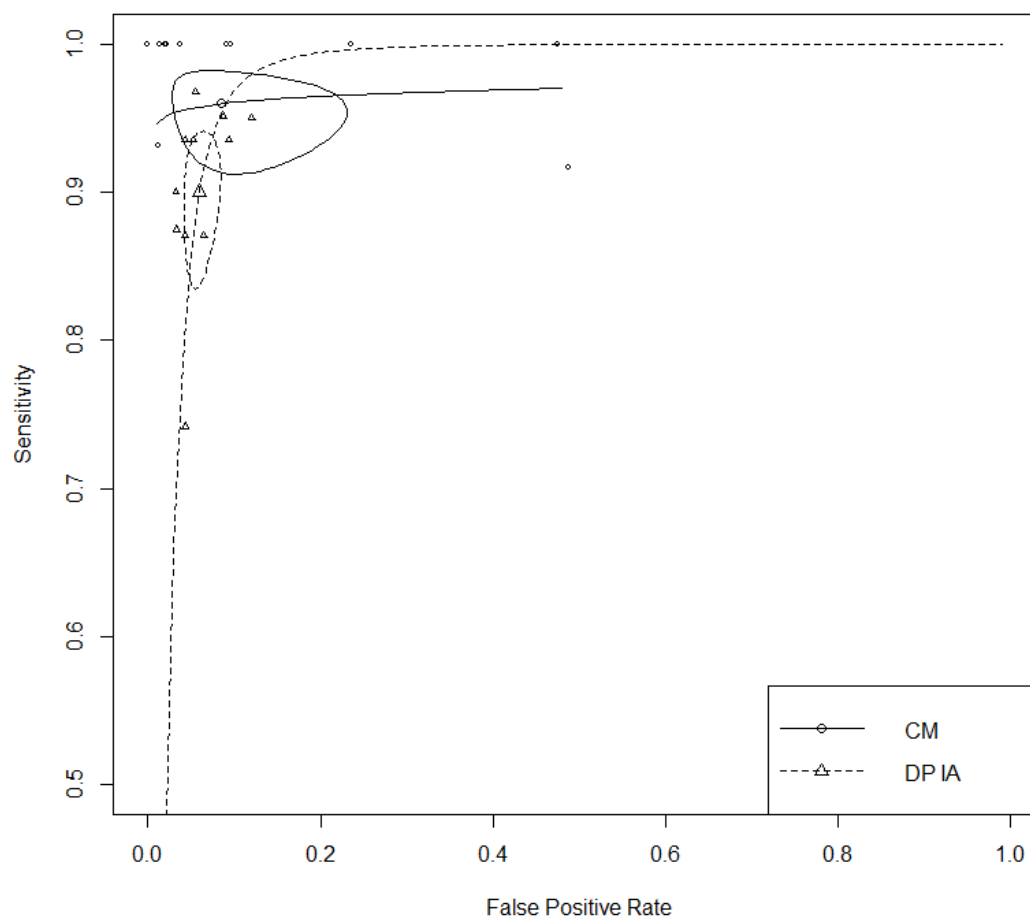


Figure 6-17 SA2 Proportion of equivocal cases if HER2 positive: CM

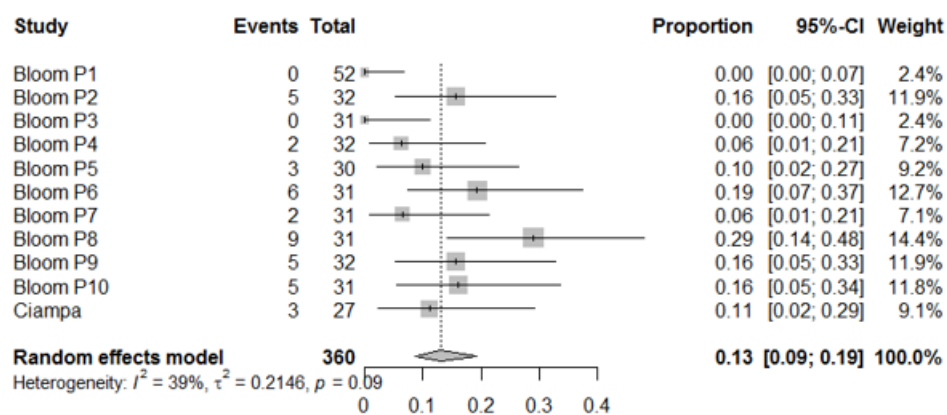


Figure 6-18 SA2 Proportion of equivocal cases if HER2 positive: DP IA

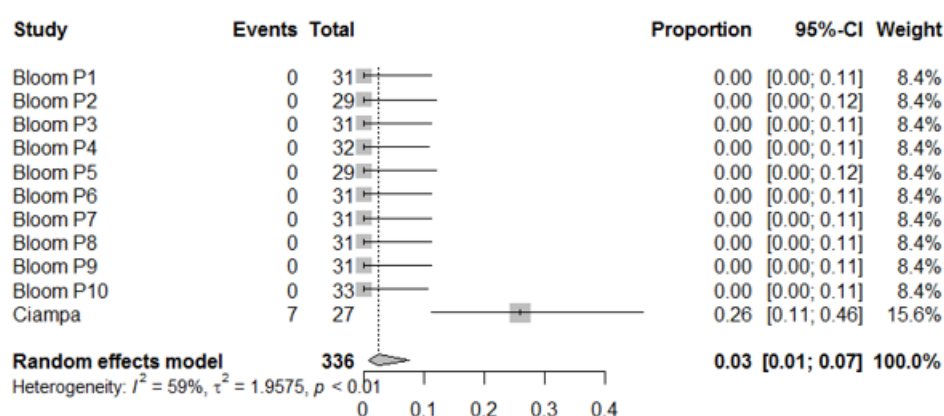


Figure 6-19 SA2 Proportion of equivocal cases if HER2 negative: CM

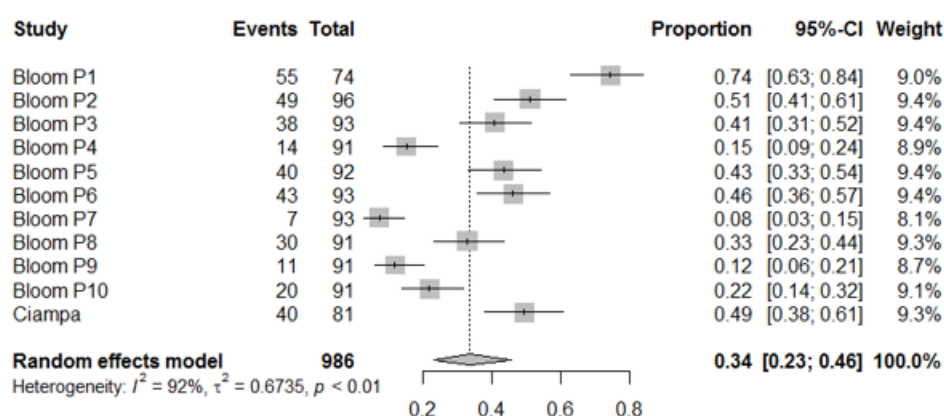
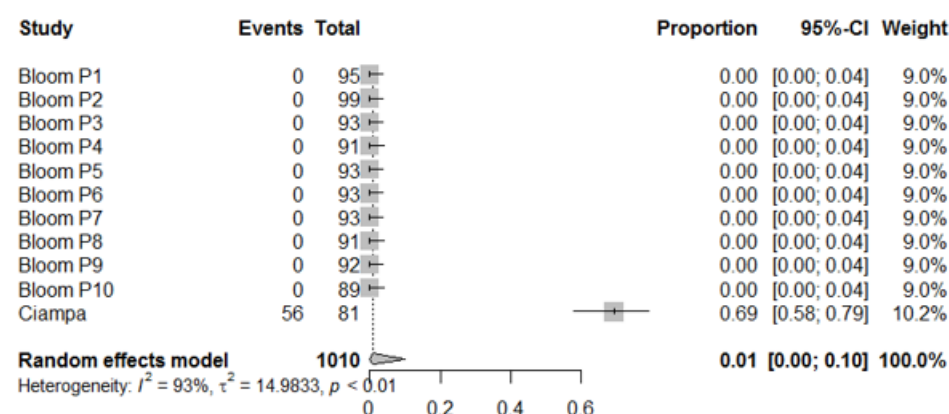


Figure 6-20 SA2 Proportion of equivocal cases if HER2 negative: DP IA



6.4.5.3 Grouping by DP IA technology

The final sensitivity analysis scenario focused on the differences in sensitivity and specificity between the DP IA technologies. The eligible studies were grouped and analysed according to the technology used as follows:

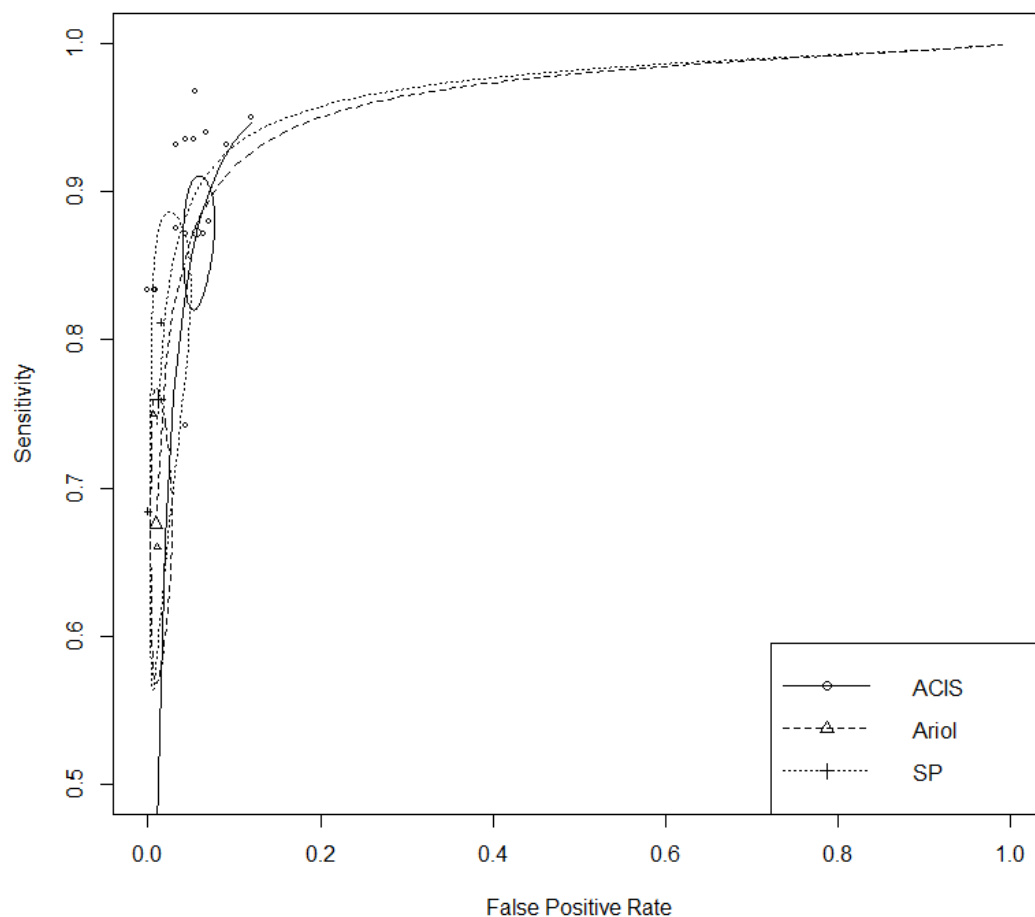
1. ACIS used by Bloom and Harrington (2004), Ciampa et al. (2006), Minot et al. (2012) and Wang et al. (2001)
2. Ariol used by Skaland et al. (2008) and Turashvili et al. (2009)
3. SlidePath used by Dobson et al. (2010) and Mohammed et al. (2012)

The remaining three studies identified through the systematic review could not be grouped and were excluded from this analysis. Table 6-8 provides a summary of the estimated sensitivity and specificity with confidence intervals for each DP IA technology and figure 6-21 illustrates the SROC curve. There is clear overlap between the confidence regions of Ariol and SlidePath, ACIS appears to have reduced specificity and possibly better able to score HER2 positive cases owing to a greater level of sensitivity compared to the other two technologies.

Table 6-8 Sensitivity and specificity estimates by DP IA technology

DP IA technology	Sensitivity	False Positive Rate
ACIS	0.873 (95%CI: 0.831, 0.905)	0.057 (95%CI: 0.043, 0.075)
Ariol	0.676 (95%CI: 0.590, 0.751)	0.009 (95%CI: 0.004, 0.022)
SlidePath	0.760 (95%CI: 0.607, 0.867)	0.011 (95%CI: 0.003, 0.037)

Figure 6-21 Comparison of DP IA technologies with SROC curves



6.5 Discussion

Accurate HER2 testing and reporting is imperative to breast cancer treatment planning. Inaccuracies result in missed or over treatment in turn impacting vulnerable patients; in particular for FN cases, preventing patients benefiting from the improvement in both overall and disease free survival seen with Herceptin treatment (Zardavas et al., 2013) and increasing the risk of cardiotoxicity for FP cases (Onitilo et al., 2014).

Any innovation in pathology designed to contribute towards a more accurate HER2 test result and so a more precise management of disease is welcome. However, it is important to be able to analyse and collate available evidence in order to make a clear judgement on the superiority of one technique over another and thus support its implementation. The systematic review and meta-analysis were designed to achieve this by estimating the sensitivity, specificity and the proportion of 2+ scores when a pathologist uses DP IA versus CM to score the HER2 IHC test.

Owing to the challenges of pooling data from studies that use different scoring thresholds, the eligible studies identified through the systematic review were grouped according to the threshold used. The base case analysis encompassed studies that followed current recommendations. The point estimates suggest a decrease in sensitivity with a concurrent increase in specificity when DP IA is compared to CM. The confidence regions for both scoring methods overlap in the ROC space, with that for digital lying nearly completely within that for CM. Whilst it is not possible to firmly conclude that HER2 accuracy using digital systems is superior to conventional methods as the overlap in confidence regions suggests no significant difference; the narrower confidence intervals for DP IA suggest it to be a more precise prediction method for HER2 scoring than a microscope. Most likely due to the reduction of variability that comes with subjective scoring when using a CM. In essence this shows that DP IA has the potential to support pathologists scoring reducing inter- and intra-observer variability. In addition, there was a statistically significant reduction in the proportion of equivocal scores across both the HER2 negative and positive groups however high levels of heterogeneity exist around these estimations.

Equivocal score proportions, sensitivity and specificity cannot be considered independently. The base case scenario suggests a more precise but not an improved estimation of sensitivity and specificity with DP IA. So while the reduction in equivocal scores is welcome there is a risk here of an increase in FP and FN cases as there is no evidence of an explicit improvement in sensitivity and/or specificity with DP IA. Reducing equivocal scores saves costs and

improves the time to diagnosis but there will potentially be a greater chance of patients receiving incorrect treatment owing to fewer patients receiving FISH testing, the gold standard in HER2 status confirmation.

Three scenarios were explored through sensitivity analysis, the first summarised accuracy estimates that were based on studies that follow the 2008 scoring thresholds to compare these to the base case scenario that adopt the 2013 thresholds. The estimates under this scenario suggest an increase in sensitivity and specificity however the confidence regions are virtually identical in ROC space for DP IA and CM. There is a reduction in equivocal scores for both the negative and positive HER2 group however these are not statistically significant due to overlapping confidence intervals. At the 2008 thresholds, the evidence suggests that there is no difference between DP IA and CM for scoring HER2 IHC.

The motivation behind the update of thresholds is the enhancement of the test's ability to distinguish between HER2 positive and negative cases. When comparing the bivariate meta-analysis outputs of the 2008 and 2013 thresholds there is a clear increase in sensitivity and a decrease in specificity across both the CM and DP IA arms when adopting the 2013 thresholds. The ability of both tests to identify positive cases has improved, decreasing the number of FN cases. There is a decrease in the ability of the IHC tests to identify negative cases so that the number of FPs have increased. With the revision of guidelines from the 2008 to 2013 thresholds, patients are more likely to be prescribed Herceptin however there is a greater number of people that are being prescribed Herceptin but don't need it. These results support the findings of other studies that have investigated the impact of changing thresholds on the given HER2 score. Stoss et al. (2015) re-evaluated the HER2 status of over 6000 breast cancer cases that were part of the HERA trial, with the updated thresholds they also reported an increase in the positivity rate and a decrease in the number of cases scored negatively.

The second scenario compared accuracy estimates for studies that matched on the thresholds used in the conventional arm but also on the technology used in the digital arm. Across the eligible studies there was a lot of variation in the technologies used for the DP IA arm of the analyses contrasting to the IHC and FISH assays which were largely consistent. DP IA systems are novel technologies which are in most cases under development and are yet to achieve approval for use. The ideal scenario for meta-analysis is to group studies on both thresholds and technology used. Two studies that met the base case criteria also matched on technology. The sensitivity and specificity outputs as well as the summary proportion of

equivocal scores were similar to those under BCA and the same trends were observed i.e. decreased sensitivity, increased specificity and reduced equivocal scores.

A final scenario explored changes in sensitivity and specificity estimates with digital technology. Results from the ACIS studies overlap with the base case DP IA results whereas SlidePath and Ariol are statistically different to base case but overlap with each other, for both of these technologies sensitivity is lower and specificity is higher than that estimated for DP IA under the base case scenario. ACIS is better able to identify positive cases whilst Ariol and SlidePath are better at identifying negative cases. When compared to CM base case accuracy, with SlidePath and Ariol, sensitivity is reduced and specificity increases. Preferably these should be evaluated against CM accuracy estimations generated by pooling results from the same studies that Ariol and SlidePath DP IA accuracy estimates are based on. However this proved challenging due to the different thresholds used in each group, limiting the number of studies that match on threshold and technology.

It is worth noting that the Bloom and Harrington (2004) study is included in both the BCA, the ideal scenario analysis and the ACIS group. This paper generated 10 pairs of 2x2 tables as ten different pathologists were involved each scoring the breast cancer slides independently. In both the base case and the ACIS group data from Bloom and Harrington (2004) study accounts for over two thirds of the included evidence. Under the ideal scenario analysis it makes up more than 90%.

Thus far the focus has been on estimating the accuracy of HER2 reporting using meta-analyses, but there are limitations to this approach. The systematic review identified in total eleven eligible studies, of these only four were included in the base case 2013 meta-analysis estimations and two in the 2008 meta-analysis.

Once the scoring thresholds have been taken into consideration there are not many published head to head studies to be included in each scenario of the review. There are 3 possible sources of heterogeneity that need to be controlled for:

1. IHC threshold in the CM arm
2. FISH threshold in the CM and DP IA arm
3. Technology used in the DP IA arm

To control for the first two excludes the majority of published head-to-head studies from the various meta-analyses. To go a step further and sub-group according to technology used significantly reduces the number of eligible studies. The limited number of studies identified,

coupled with the wide range of technologies and algorithms used, each with their own unique HER2 scoring methods introduces a degree of heterogeneity into both the 2013 and the 2008 group.

Across the 11 studies included in the review, 7 different digital technology systems are used for HER2 scoring. These will differ in a number of ways including; the components they are made up of, the illumination method, computer hardware, monitors used, the digitisation process and digital slide visualisation techniques (Rojo et al., 2006). Keay et al. (2013) investigated the reproducibility of HER2 scoring across 3 different digitisation systems and across two different image analysis algorithms. There were high levels of variability in the results between systems and also between algorithms (Keay et al., 2013). However a review of the published evidence on reproducibility found levels to be acceptable and in no way poorer than that found when using conventional approaches (RIBER-HANSEN et al., 2012). Bloom and Harrington (2004) found inter-observer agreement among the 10 pathologists to improve with digitisation however each pathologist used identical technology for their DP IA scoring. Further work on the standardisation of digital techniques and algorithms needs to be carried out to limit inter-laboratory variability and ensure all HER2 cases are scored appropriately especially as there is some evidence of reduced equivocal scores with digitisation.

The proportion of equivocal scores varied greatly between the head-to-head studies included in the review; ranging from 5 - 44% for conventional scoring and 0 - 58% for digital technology. Revised ASCO/CAP guidelines recently released in July 2018 refine the definition of 2+ scores (Wolff et al., 2018). As there has always been confusion on the classification of equivocal HER2 cases; pathologists are cautious, giving a 2+ score where there is uncertainty (Helin et al., 2016) to ensure the gold standard more definitive FISH test is carried out. This is more evident among junior rather than senior pathologists. With the updated guidelines, it is anticipated there will be greater uniformity in scoring this indeterminate HER2 group. If the revised guidelines are able to bring CM scoring results closer to the gold standard outcomes, arguably DP IA accuracy maybe equivalent to subjective pathologist scoring when guidelines are in place and applied consistently. It is too early to know if this will be the case and further research is needed as the new guidelines are implemented across laboratories.

Equivocal test results can essentially be classified as a non-evaluable index test result since they are neither negative nor positive and have simply been omitted from the bivariate meta-analysis when constructing the 2x2 tables. Questions arise as to whether this is the

best way to handle such results when carrying out a meta-analysis. Schuetz et al. (2012) compared diagnostic accuracy outcomes when non-evaluable results are omitted or all classified positive or all classified as negative. They found both sensitivity and specificity to be overestimated under the exclusion method and one or the other to be inflated when all non-evaluable results are classed as negative or positive. The European network for health technology assessment recommend in their guidelines on meta-analysis of diagnostic test accuracy the use of an intention-to-diagnose approach and include all non-evaluable results in 3x2 tables (EUnetHTA, 2014). Equivocal results from patients that have the disease are considered false negative and equivocal results of patients that do not have the disease are considered false positive (Ma et al., 2014). Unfortunately adopting this method would not meet the review objectives; the sensitivity and specificity are not the only outcomes of interest but also independently measuring the probability of obtaining an equivocal score needed to be established. A more appropriate method would be to extend the bivariate model to take account of the equivocal risk by using a trivariate model as demonstrated by Chu et al. (2009). Sensitivity, specificity and equivocal proportions can be simultaneously analysed to take account of correlation between each effect (Riley et al., 2017). A trivariate model will determine whether the sensitivities and specificities worked out from the included studies are dependent on the probability of obtaining a definite score on the first round of testing (Chu et al., 2009).

The review could have also go further in its analysis by taking into consideration the level of experience of the reporting clinician in each primary study. DP IA is a means to overcome discrepancies in subjective scoring that may result from differences in levels of experience (Walker, 2006). For example at the individual study level the scoring fluctuations observed between CM and DP IA are perhaps greater with pathologists at the start of their career in comparison to more experienced consultants. However to take this into account would have undoubtedly meant a reduction in the overall number of included studies making it impractical to carry out a meta-analysis of this nature. For the same reasons it was also not possible to separate out and analyse the primary research by level of quality as there were not enough studies of similar quality that use identical thresholds to score HER2.

There is much debate on the most appropriate study design method for estimating diagnostic test accuracy. An alternative approach to cohort studies is the use of RCTs. A large well conducted RCT will overcome the issues of study heterogeneity and differences in population (Walker et al., 2008) found when carrying out a meta-analysis of much smaller cohort studies. While RCTs are considered the gold standard in intervention studies they are

extremely rare when investigating diagnostic accuracy (Rodger et al., 2012), none were picked up through the systematic review process. Beyond meta-analyses and RCTs there is potential to make accuracy estimations based on a single good quality cohort study. Diagnostic studies involve the management of glass slides rather than individuals and all tests of interest are generally conducted on all tissue specimens in the study sample; no doubt making them less prone to the risks of bias experienced with intervention studies. Certain biases for example performance, detection and attrition (Mansournia et al., 2017) are not applicable.

6.6 Conclusion

While there is no robust evidence and further primary research is needed to support the superiority of digital systems over conventional microscopes there is also no indication that it is inferior and in reality it is expected to be at least equivalent to a pathologist's subjective scoring. Nonetheless the benefits of DP IA in HER2 breast cancer must be measured against the cost implications of the adoption of such a system in pathology. This information can be used in economic evaluations to determine the cost-effectiveness of DP IA in scoring HER2 IHC and results used to support the adoption or rejection of this technology.

Chapter 7: Economic evaluation of digital HER2 testing

7.1 Introduction

A systematic review and meta-analyses comparing HER2 IHC test accuracy when using CM and DP IA were carried out and reported in the previous chapter. Meta-analyses were used to pool data and to generate test accuracy estimates for CM and DP IA to score HER2 IHC, with the results demonstrating improvements in the precision of scoring with digital methods and a reduction in the number of equivocal cases. Evidence of accuracy changes with digitisation are of limited benefit to decision makers if not considered in light of their implications for clinical and cost outcomes to support adoption.

7.1.1 Patient care

Concurrent to weighing up accuracy developments with new diagnostics, attention must be given to the grander scheme of patient care. Test accuracy alone does not improve clinical outcomes; yet it is the pivot on which investment and adoption decisions are made (Rodger et al., 2012). It is necessary to include breast cancer treatment following testing to assess and value the full benefit of scoring HER2 IHC using digital methods. Herceptin treatment, a clinical process that sits between the HER2 biomarker and the patient, its prescription or avoidance as a result of the test's outcome defines whether any clinical or cost benefits are seen. An ideal scenario would feature reductions in FP and FN cases and an increase in diagnoses that match to the patient's true HER2 status with digitisation, leading to greater precision in the design of care plans and limiting avoidable NHS costs. Diagnostic RCTs or modelling approaches where both the test and the therapeutic intervention are combined in one study (Lu and Gatsonis, 2013) can be used to determine the value to breast cancer care of DP IA to score HER2 IHC.

7.1.2 Economic evaluation

Economic evaluations can be used to predict changes in the cost and patient outcomes with the introduction of new DP IA technology for HER2 IHC testing. By taking a modelling approach the appropriate connection can be made between the HER2 testing algorithm, test performance under each intervention, treatment prescribed and patient and cost outcomes. Each of the IHC testing methods can be represented by manipulating model parameters to reflect each method appropriately. The main adaptation in the HER2 model from CM when evaluating DP IA are the probability values that are input since these are largely driven by the test performance estimated under each method.

Economic evaluation results can be expressed in terms of incremental net benefit (INB) where both costs and effects (e.g. QALYs) are valued and expressed in monetary terms, using

the NICE recommended willingness to pay threshold of £20,000 to £30,000 (McCabe et al., 2008). Any INB result above zero suggests that the intervention is cost effective and supports its adoption. For this to occur the monetary value of the difference in effect must be greater than the difference in costs between conventional and digital. Conversely, a negative INB does not provide evidence to support the adoption of digital techniques and based on the economic evaluation alone the concept of pathology digitisation will be rejected.

No previous economic evaluation has been carried out to justify the introduction of digital pathology image analysis into the UK Healthcare setting. The results of the economic evaluation can be used to inform decision making relating to digital pathology while recognising that the effects of digitisation will undoubtedly be realised beyond this single biomarker. Digitally upgrading any pathology lab will not only be of value when scoring HER2 or even just in the management of breast cancer cases; but rather it will replace every task in the pathology lab where a microscope is used. Ideally the introduction of digital techniques should be weighed up against possible accuracy benefits across all disease areas but in reality this is challenging due to the extent and range of available biomarkers and tests where digital technology can theoretically have an influence. For this reason it is difficult to explore the cost-effectiveness of digitisation across all tests of interest but only from the perspective of one biomarker at a time.

All of the eligible studies identified in the systematic review of economic evaluations that was carried out in chapter 2 used estimates of sensitivity and specificity based on single studies. None attempted to pool estimates of test performance using meta-analysis techniques to inform their model. This chapter also explore the impacts on cost-effectiveness estimates when accuracy parameters are based on a good-quality single study compared to when they are based on pooled estimate

7.2 Aims

This economic evaluation will focus on the impact of digital pathology image analysis when scoring HER2 IHC as part of invasive breast cancer management process.

1. The primary aim is to determine the cost-effectiveness of Digital Pathology Image Analysis when compared to Conventional Microscopes in determining HER2 status for invasive breast cancer patients.
2. The secondary aim of the study is to explore changes in decisions made when HER2 accuracy estimates are based on one good quality cohort study compared to when they are based on the results of a meta-analysis

7.3 Methods

The cost effectiveness analysis was designed to measure the change in costs and QALYs for invasive breast cancer patients with the introduction of digital pathology image analysis technology for scoring HER2 IHC compared to conventional microscope methods. The analysis involved several stages. Firstly, a conceptual model was developed, this was simply extracted from the information learnt about the HER2 testing algorithm that had previously been incorporated into the conceptual model of the DES model described in chapter 5. The conceptual model allowed the identification of all the model parameters; estimates of these were taken from various sources including both single and synthesis based studies. The model structure and its parameters were computed using R software and designed to generate an INB result. Sensitivity analyses were carried out to establish the impact of parameter uncertainty on the INB estimations. In addition a value of information analysis (VOI) was performed to measure the expected gain from further research.

7.3.1 Model structure

A decision tree model was developed to be used for the economic evaluation; it principally aimed to capture all the stages of the HER2 testing strategy. The tree structure has two arms with identical branches; one representing digital pathology image analysis and the second conventional microscopes. A schematic of one of the arms is given in figure 7-1; the model structure for the digital pathology image analysis arm is a replica of that for conventional microscope examination. To start a decision node divides the tree into each arm of the analysis. At this point based on UK HER2 prevalence data; patients are designated to be either HER2 positive or negative reflecting their true status. Of these a fraction are equivocal prior to HER2 status confirmation by FISH testing and as such are directed to the Equivocal (EQ) arm. Those that are Not Equivocal (NEQ) initially on IHC and are given a HER2 result are directed to the NEQ arm. If their true HER2 status is positive; they can either be a True Positive (TP) or a False Negative (FN) on IHC. If their status is negative; they are split between those who are True Negative (TN) or False Positive (FP).

Each endpoint in the tree represents a unique combination of events for invasive breast cancer patients; these are the true HER2 status, the final result given and the consequent treatment plan. Table 7-1 describes these. There are 6 possible scenarios regarding a patient's HER2 pathway. For two of these (C and F); additional FISH testing is required, so while they are ultimately given the correct treatment, the additional test causes delay and adds an extra cost to the pathway. For scenarios B and E incorrect treatment is prescribed; one where it is omitted (B) and the second where it is given in error (E). The introduction of

DP IA is anticipated to reduce the probability of these four combinations of events and increase the likelihood of the remaining ideal two scenarios (A and D) where breast cancer patients are given the correct diagnosis on the first round of testing and thus prescribed Herceptin appropriately.

Figure 7-1 Illustration of the convention microscope arm of the decision tree

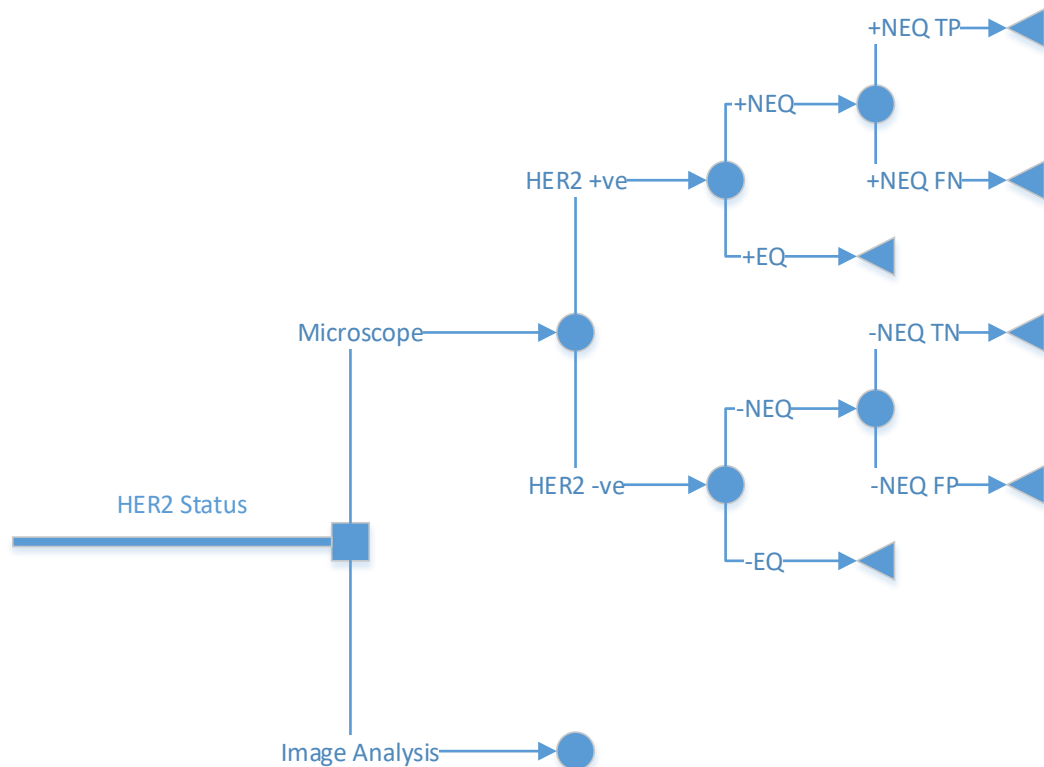


Table 7-1 Decision tree endpoints

	End Point	True Status	IHC	FISH	Herceptin	Care Plan
A	+NEQ TP	Positive	+3	----	Yes	Correct
B	+NEQ FN	Positive	0/+1	----	No	Incorrect
C	+EQ	Positive	+2	Amplified	Yes	Correct
D	-NEQ TN	Negative	0/+1	----	No	Correct
E	-NEQ FP	Negative	+3	----	Yes	Incorrect
F	-EQ	Negative	+2	Non-Amplified	No	Correct

Probabilities are included following each chance node in the tree as well as costs and QALYs are added to the end points of each branch. While each arm of the decision tree model is identical on its structure, QALYs and costs, the attached probabilities differ between the two arms. Cost and QALY differences between the CM and DP IA were valued over the breast

cancer patient's lifetime by using total lifetime QALY estimates per patient for Herceptin treatment estimated and reported by Clarke et al. (2017).

The model assumes that all patients will receive a conclusive HER2 score by the end of the two-step testing process; in reality this is not always the case. For various reasons such as inadequate specimen submitted; a definite HER2 result may not be reached even post FISH and thus generally for these patients a recommendation is made to repeat the HER2 testing process on a second specimen. The model also assumes that all patients are treated with Herceptin if they are found to be HER2 positive; in reality this is also not the case. If age, co morbidities or patient preference are taken into account a fraction of breast cancers will not be treated with Herceptin even though they are HER2 positive. The model also does not differentiate between primary and metastatic cases and considers all invasive breast cancers a homogenous group.

7.3.2 Model costs

Costs are assigned at the endpoints of each branch of the decision tree. The cost effectiveness analysis took an NHS Perspective and as such only included costs to the NHS directly related to HER2 testing and Herceptin treatment process. For each branch a combination of costs are incorporated; these are the cost of (1) IHC Test (2) FISH Test, including both the cost of the test itself and delivery and (3) the cost of Herceptin Treatment. The decision tree model follows UHCW HER2 testing processes and current UK recommendations as outlined in figure 6-2 in chapter 6 and assumes all FISH testing is carried out off-site. The reference cost year was 2015-2016; adjusted using inflation indices from PSSRU (2016) where data was not available for this period.

The cost of basic IHC slide preparation is the same for both the digital and the conventional arm and so this is not included as an input in the model. A business case was put together at UHCW to support the introduction of the digital system into their pathology labs and for its use across the trust. As part of this, the projected costs of digital pathology over a 7-year period were detailed, this does not include the cost of the HER2 image analysis algorithm. Based on UHCW data the annual figures for slides requested over the 7-year period could also be extrapolated. All histopathology slides were included since in reality digital pathology will be used across all diseases not just breast cancer cases. Based on this information, the system installation will cost an extra £0.64/slide. Due to the high throughput and consequent minimal cost/slide, the cost of digital pathology and image analysis were not included in the model at this stage of the analysis. The focus will be on measuring the changes in treatment costs and benefits as a result of accuracy implications with DP IA.

The cost of FISH testing is based on UHCW prices; this is currently £115 and includes both the cost of test and delivery. For the analysis, the Herceptin treatment duration is assumed to be 9 weeks and sensitivity analysis will explore the impact on INB if this was at the currently recommended 12 month treatment duration. Herceptin costs for both 9 weeks and 12 months treatment are based on values from Clarke et al. (2017); their 2013/2014 costs are adjusted to 2015/2016 costs using PSSRU (2016), these are expressed and input into the model as normal distributions. The costs are identical for both arms of the decision tree since the cost of digital pathology image analysis is not included at this stage of the economic evaluation. Example mean cost values for each branch of the tree are given in table 7-3; these are based on the base case 9 weeks Herceptin treatment.

Table 7-2 Cost of Herceptin treatment (Clarke et al., 2017)

Duration	Distribution	Parameters
9weeks	Normal	£7666 (95% CI: £6,831, £8,501)
12months	Normal	£28,494 (95% CI: £25,064, £31,924)

Table 7-3 Example 9 week mean model costs input

Branch	Included Costs	Total
+NEQ TP	Herceptin	£7666
+NEQ FN	None	£0
+EQ	Herceptin + FISH	£7781
-NEQ TN	None	£0
-NEQ FP	Herceptin	£7666
-EQ	FISH	£115

7.3.3 Model QALYs

In addition to costs; QALY values are assigned to the various decision tree endpoints. For the TP and FN arms; these are taken from Clarke et al. (2017). Their cost effectiveness analysis undertook a network meta-analysis where all trials recruited patients who were HER2 positive. The outcomes seen in the observation; zero Herceptin arm are assumed to be equivalent to those breast cancer cases with HER2 positive status but not given treatment due to error in reporting results i.e. they are FN. The consequent QALY outcomes for the 9week and 12month treatment arms are used for TP status. TP and FN QALYs are expressed and input into the model as normal distributions, these are given in table 7-4.

The QALYS for the TN and FP arms were estimated as these values were not available in the literature. Katzorke et al. (2013) compared outcomes in HER2 positive breast cancers treated with Herceptin to those HER2 negative cases. Overall survival is greater in the former compared to the latter; with a hazard ratio of 0.67 (95% CI: 0.48, 0.92) (Katzorke et al., 2013). The mean hazard value is used to calculate the TN QALY from the TP QALY distribution.

FP patients are those whose true status is negative but receive Herceptin due to reporting errors; their QALY values are calculated by subtracting the disutility of 0.104 (95%CI: 0.084, 0.0125) (Garrison et al., 2015) associated with cardio toxicity experienced when a patient is given Herceptin treatment. As with costs; base cases analysis was based on the 9 week treatment period and sensitivity analysis will explore changes in INB when 12 months of Herceptin is prescribed.

Table 7-4 QALYs of Herceptin treatment

Arm	Distribution	Parameters	
		9 weeks	12months
TP	Normal	10 (95%CI: 8.62, 11.38)	9.2 (95%CI: 7.95, 10.45)
FN	Normal	8.6 (95%CI: 7.45, 9.75)	8.6 (95%CI: 7.45, 9.75)

For the +EQ branch the QALY will be identical to TP values and for the –EQ branch the QALY input will be equal to the TN values. The QALY values are the same for both the digital and conventional arms of the decision tree.

7.3.4 UK prevalence

The incidence of HER2 positivity is based on a 5 year UK national audit estimating the overall HER2 positivity in invasive breast cancer patients at 14.5% (Rakha et al., 2014).

7.3.5 Model probabilities

The test effect was measured using odds ratios (OR). Estimates of the DP IA arm probabilities were derived by application of Log(OR) for the test effects to CM arm probabilities reported in the published literature as illustrated in the equation below:

$$\text{logit}(p_{2x}) = \text{logit}(p_{1x}) + \text{Log}(\text{OR}_x)$$

Welton et al. (2012)

p1 are the probabilities used to populate the CM arm of the model, p2 are the probabilities for the DP IA arm and x defines the probability of interest as listed, where x can be one of the following:

1. +eq the probability of an equivocal score if HER2 positive
2. -eq the probability of an equivocal score if HER2 negative
3. Sens the probability of a TP result in the positive non-equivocal group
4. FPR the probability of a FP result in the negative non-equivocal group

There are four test effects to be estimated on the log scale to be applied in the above equation, these are listed in table 7-5.

Table 7-5 List of test effects

Log(OR _x)	Description
Log(OR _{+eq})	Odds of an equivocal result if HER2 status is positive in the DP IA arm relative to the odds of an equivocal result in the CM arm
Log(OR _{-eq})	Odds of an equivocal result if HER2 status is negative in the DP IA arm relative to the odds of an equivocal result in the CM arm
Log(OR _{sens})	Odds of a TP result in the DP IA arm relative to the odds of a TP result in the CM arm
Log(OR _{FPR})	Odds of a FP result in the DP IA arm relative to the odds of a FP result in the CM arm

Distributions for p_{1x} and Log(OR_x) differ between base case and the sensitivity analysis scenarios. These are described under each.

7.3.6 Base case scenario

This analysis estimated p_{1x} and log(OR_x) from studies that met the criteria for inclusion in the systematic review described in chapter six and adopted the 2013 thresholds for scoring. The OR for each of the four test effects from each study were pooled on the log scale using the *metafor* package in R, the mean and standard error (Se) of the Log(OR) are given in table 7-6.

Table 7-6 Estimates of Log(OR)

Log(OR _x)	2013 data		2008 data		2013 single study data	
	mean	Se	mean	Se	mean	Se
Log(OR _{+eq})	-1.1041	0.4487	-0.1199	0.3489	-0.5994	0.1771
Log(OR _{-eq})	-2.9553	0.7421	-0.4054	0.1921	0.1852	0.3725
Log(OR _{sens})	-0.8079	0.3692	0.0173	0.4766	-0.3127	0.2248
Log(OR _{FPR})	-0.3453	0.4719	-0.0601	0.8194	0.3453	1.5111

P1 probabilities are either based on the meta-analyses carried out in chapter 6 or are based on data extracted directly from the published literature. In the base case analysis p_{1+eq} and p_{1-eq} are from Mohammed et al. (2012), this is a good quality head-head study that adhered to the 2013 thresholds for scoring. p_{1+eq} and p_{1-eq} were expressed and input into the model as beta distributions in the CM arm of the model. The same distributions were transformed onto the logit scale and the $\log(OR_{+eq})$ applied to $\text{logit}(p_{1+eq})$ and $\log(OR_{-eq})$ to $\text{logit}(p_{1-eq})$ to estimate $\text{logit}(p_{2+eq})$ and $\text{logit}(p_{2-eq})$ respectively. The resultant distributions were back-transformed to generate the probability distributions for p_{2+eq} and p_{2-eq} to be input as distributions into the DP IA arm of the model.

p_{1sens} and p_{1FPR} were taken from the outputs of base case bivariate meta-analysis in chapter 6. These were already expressed on the logit scale, $\log(OR_{sens})$ and $\log(OR_{FPR})$ were applied to these to generate $\text{logit}(p_{2sens})$ and $\text{logit}(p_{2FPR})$. All logit probabilities were back transformed and used to populate the model. Summary of the p1 probabilities that are used to populate the CM arm of the model are given in table 7-7.

Table 7-7 BCA p1 distributions

p1	Distribution	Parameters	Source
p_{1+eq}	Beta	$\alpha=18 \beta=62$	Mohammed et al. (2012)
p_{1-eq}	Beta	$\alpha=5 \beta=346$	
$\text{logit}(p_{1sens})$	Normal	2.652 (95% CI: 2.238, 3.066)	Chapter 6: Bivariate meta-analysis results
$\text{logit}(p_{1FPR})$	Normal	-2.735 (95% CI: -3.167, -2.303)	

The Costs and QALYS associated with the 9-week treatment period are used in the base case analysis due to its dominance over 12 month treatment.

A deterministic version of the model is initially run by applying point estimates of the mean for all the model parameters, this is followed by probabilistic version of the model is run where all parameter inputs are based on the distributions outlined previously to generate a probability for cost-effectiveness.

To further explore cost-effectiveness implications of DP IA and ascertain the robustness of decisions made, four further scenarios are explored as described below by making appropriate modifications to the model. Scenario analyses are all carried out using stochastic models to allow for random variation.

7.3.7 Scenario analysis 1: 2008 scoring thresholds

Under this scenario the impacts on cost-effectiveness when the test effect estimates are based on the 2008 scoring thresholds are explored. Identical cost and QALY distributions to those in the BCA were used to populate the model. The model structure and methods for assessment of p_{2x} are also identical. Rather than using data from the 2013 threshold studies, all $\log(\text{OR}_x)$ are based on the 2008 threshold studies that were eligible for inclusion in the systematic review in chapter 6. Summary of the distributions are given in table 7-8.

For p_{1x} distributions, the same approach was taken as that for the base case. p_{1+eq} and p_{1-eq} were based on Minot et al. (2009), a good quality head-to-head study that adopts the 2008 thresholds for scoring. For estimations of p_{1sens} and p_{1FPR} the outputs from the bivariate meta-analysis in chapter 6 for the 2008 studies was used. A summary of the p_1 distributions as they are input into the model is given in table 7-8.

Table 7-8 SA1 p_1 distributions

p_1	Distribution	Parameters	Source
p_{1+eq}	Beta	$\alpha=7$ $\beta=10$	Minot et al. (2009)
p_{1-eq}	Beta	$\alpha=24$ $\beta=113$	
$\text{logit}(p_{1sens})$	Normal	1.064 (95% CI: 0.75, 1.378)	Chapter 6: Bivariate meta-analysis results
$\text{logit}(p_{1FPR})$	Normal	-4.507 (95% CI: -5.01, -4.004)	

7.3.8 Scenario analysis 2: 12 month treatment period

Whilst published studies have shown the 9-week treatment period for Herceptin to dominate 12 months, the latter still remains the recommended treatment length in the UK. This scenario explores cost-effectiveness outcomes using cost and QALYs associated with 12-month treatment. These are listed in table 7-2 and table 7-4. Model probabilities are identical to those used in the base case model.

7.3.9 Scenario analysis 3: Single study approach

The third scenario explores the impact on cost-effectiveness when the probabilities and test effect are estimated using a good quality head-to-head study. The systematic review in chapter 6 identified Mohammed et al. (2012) as a good quality head-to-head study; this was the only study adhering to the 2013 thresholds that scored positively on all the domains of the QUADAS-2 tool as illustrated in table 6-6. Data from this study was used to inform this scenario analysis.

QALY and cost distributions are identical to those used in base case analysis however values for $p1_x$ are expressed and input into the model as beta distributions using data extracted from Mohammed et al. (2012) and $\log(OR_x)$ are expressed as normal distributions and are also based on data extracted from Mohammed et al. (2012). $p2_x$ values are estimated using the same methods as described previously. Table 7-6 and 7-9 summarise estimates of $\log(OR_x)$ and $p1_x$ respectively.

Table 7-9 SA3 p1 distributions

p1	Distribution	Parameter	Source
$p1_{+eq}$	Beta	$\alpha=18 \beta=62$	Mohammed et al. (2012)
$p1_{-eq}$	Beta	$\alpha=5 \beta=346$	
$p1_{sens}$	Beta	$\alpha=53 \beta=9$	
$p1_{FPR}$	Beta	$\alpha=2 \beta=344$	

7.3.10 Cost-effectiveness analysis

The model equations are designed to start at the decision tree end point, applying back calculations to determine the total QALYs and costs associated with each arm of the model. The incremental cost and QALY between the DP IA and CM arms are estimated using the following equations:

$$\text{Incremental QALY} = \text{DP IA QALY} - \text{CM QALY}$$

$$\text{Incremental Cost} = \text{DP IA Cost} - \text{CM Cost}$$

50000 observations are generated from all model distributions. The model $\text{inc}(\text{cost})$ and $\text{inc}(\text{QALY})$ are plotted against each other on the cost-effectiveness plane for each scenario. These provide a visual representation of cost-effectiveness observing the distribution of the 50000 simulated results on the quadrants of the plane.

The probability of cost-effectiveness was estimated by using incremental INB values at each WTP threshold. The incremental QALY distributions generated under each scenario are converted into a distribution of monetary values through multiplication by the WTP threshold. These values are converted to INB as follows:

$$\text{INB} = (\text{Incremental QALY} \times \text{WTP threshold}) - \text{Incremental Cost}$$

A positive INB indicates that the intervention is cost-effective whereas a negative INB does not favour the intervention. 50000 simulations of INB are generated at each WTP threshold. The total number of positive estimations is divided by 50000 to generate the probability of

cost-effectiveness at each WTP threshold. Each probability was plotted against its WTP threshold to generate a cost-effectiveness acceptability curve (CEAC) demonstrating how cost-effectiveness alters with WTP for each scenario.

7.3.11 Value of Information Analysis

The value of information (VOI) analysis quantifies the value of reducing uncertainty through further research (Wilson, 2015) to inform a decision problem. Two levels of VOI were carried out for each scenario. The expected value of perfect information (EVPI) is the difference between a decision being made based on perfect information where uncertainties around all parameters are eliminated, and the decision being made based on the existing imperfect information (Claxton and Sculpher, 2006). EVPI was estimated for each scenario at both the £20,000 and the £30,000 WTP threshold. The mean net benefit at the WTP thresholds is known, the 50000 simulated INB values are each compared against the mean net benefit. If the investment decision is altered with the individual INB value compared to the decision that was made with the mean INB the individual INBs are summed and averaged, this is the EVPI. The EVPI quantifies the value of eliminating uncertainty around all the model parameters.

Further to patient level estimates, EVPI can also be quantified at the population level, this is a more appropriate value to compare to the costs of further research (Thorn et al., 2016). The incidence of invasive breast cancer cases over a suitable time horizon was multiplied by the patient EVPI to calculate a value for population EVPI. A discount rate of 3.5% was applied. Further research can be justified if the cost of investment falls below the population EVPI (Oostenbrink et al., 2008).

The EVPI value alone is not sufficient on which to base research funding decisions (Thorn et al., 2016). The expected value of partial perfect information (EVPPI), quantifies the difference in expected INB when perfect information is available for each of the model parameters. Similarly to the methods for EVPI but an additional loop is added, for the model parameter of interest the value is fixed for each run of the model that generates 50000 estimations of INB. This loop is repeated for each of the 50000 values of the parameter of interest. The decision made using the mean INB for each loop is compared against the decision made prior to EVPPI analysis, where there is a change the mean INBs are summed and averaged to generate a value for the elimination of uncertainty around the parameter of interest. Each parameter is investigated independently defining where future research should focus. EVPPI was carried out for the base case analysis and

7.4 Results

7.4.1 Base case analysis results

The decision tree model for base case analysis was initially run as a deterministic model to estimate the INB of DP IA compared to CM when scoring HER2 IHC. The INB was negative at both the £20,000 and the £30,000 WTP thresholds. A summary of these results is given in table 7-10 together with the incremental costs and QALYs when comparing DP IA to CM.

Table 7-10 Deterministic model outputs

Strategy	Model outputs		INB results	
	QALY	Cost	WTP £20000	WTP £30000
CM	7.17	£1453.85	-£108.52	-£258.62
DP IA	7.16	£1262.46		
Increment	-0.01	-£191.38		

The deterministic model outputs show DP IA as not cost-effective at both WTP thresholds because the monetary value of the QALY loss outweighs the cost-savings observed. The costs and QALY outputs of the decision tree would lie in the south-west quadrant on the cost-effectiveness plane; as digitisation is predicted to be both cost saving and QALY reducing. Due to the QALY losses experienced when moving from conventional to digital techniques, a negative value is generated in the first part of the INB calculation. An increase in the WTP gives a greater monetary value to the QALY loss incurred and so an increase in the threshold is associated with a concurrent decrease in INB.

The base case scenario model was run again with all the model parameters expressed as distributions rather than as point estimates. 50000 estimations of incremental cost and QALY were generated and are illustrated on the cost-effectiveness plane in figure 7-2. The two lines on the figure represent the NICE WTP thresholds, any points falling below these favour digitisation, contrarily points above the WTP thresholds favour conventional microscope techniques.

On review of figure 7-2 it is obvious that judgements on the cost-effectiveness of digitisation are not so clear-cut. Unlike what is seen under the deterministic analysis, digitisation does not always generate a negative INB outcome. There is both evidence to support and to counter the cost-effectiveness of digitisation. The results fall across all four quadrants however it appears that there may be a low probability of cost-effectiveness due to the position of the points in relation to the WTP thresholds with what appears to be more than 50% lying above the lines.

To explore the level of uncertainty around the cost-effectiveness, the points on the plane in figure 7-2 are manipulated to form the cost-effectiveness acceptability curves (CEAC) in figure 7-3. The probability of cost-effectiveness is plotted against a series of WTP thresholds. At the £20,000 WTP threshold there is a probability of 0.41 of cost-effectiveness and at the £30,000 WTP threshold decrease to 0.35, probability cost effectiveness is reducing with the threshold, due to the higher value attached to the QALY losses observed.

Figure 7-2 BCA cost-effectiveness plane

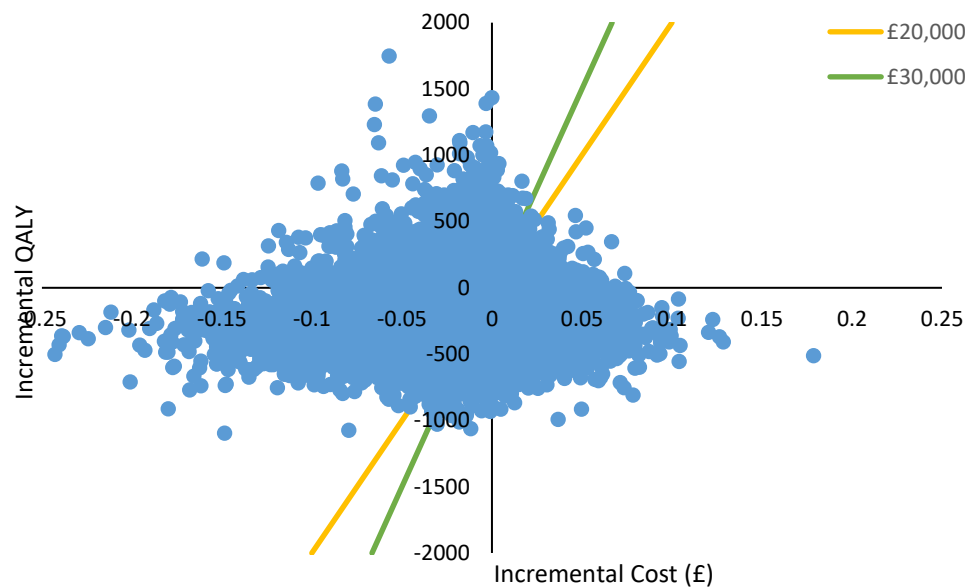
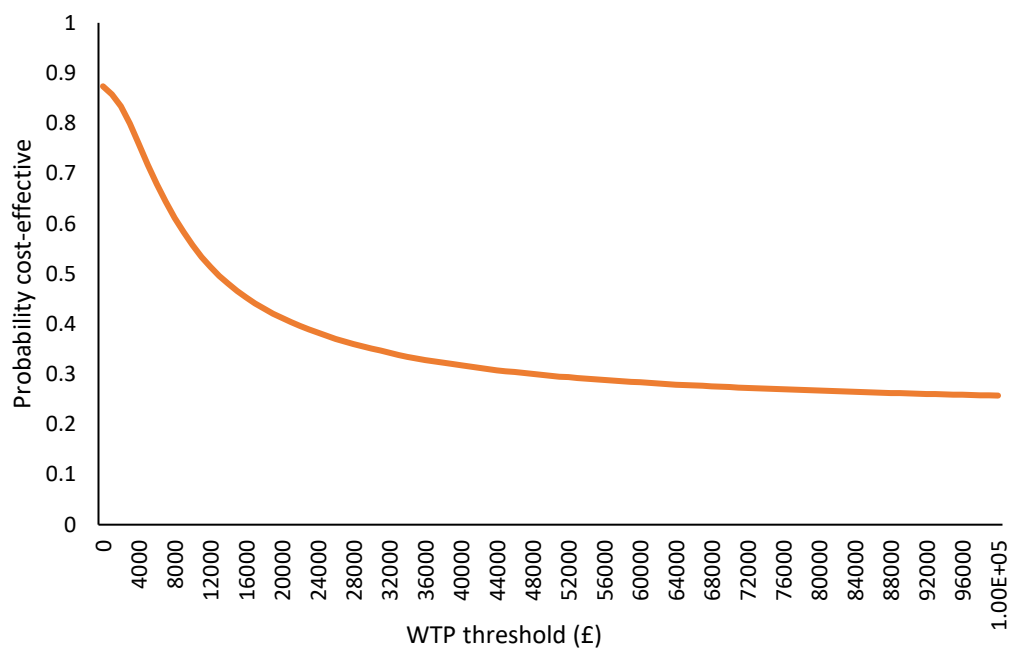


Figure 7-3 BCA CEAC



7.4.2 Scenario analysis results

Cost effectiveness planes and CEAC were generated for each scenario analysis, the cost effectiveness planes are given in figures 7-4, 7-5 and 7-6 for each scenario and the CEACs in figure 7-7.

Figure 7-4 SA1 cost-effectiveness plane

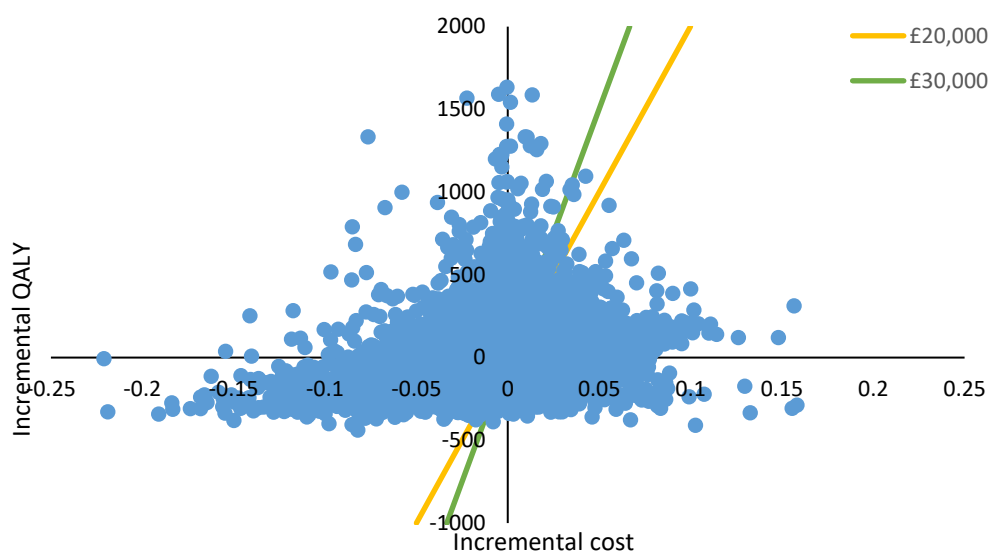
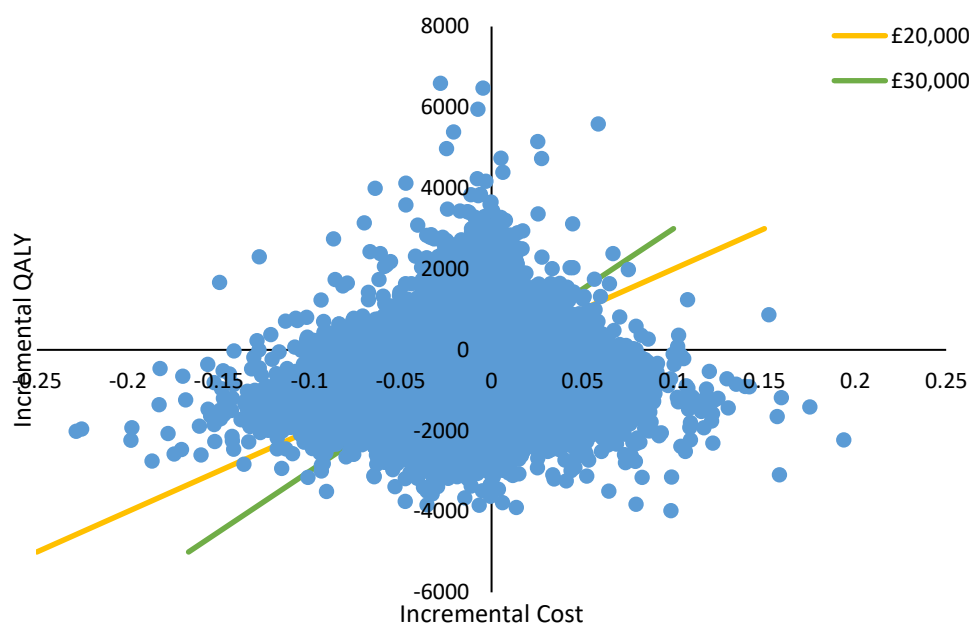


Figure 7-5 SA2 cost-effectiveness plane



The scatter of points on the cost-effectiveness planes for the SA1 scenario, where ORs and probabilities are based on studies that use 2008 scoring thresholds and for the SA2 scenario where costs and QALYs are based on 12 months of Herceptin treatment are noticeably similar

to the base case cost-effectiveness plane. For the SA3 scenario the scatter of points obviously differ from the others.

Figure 7-6 SA3 Cost-effectiveness plane

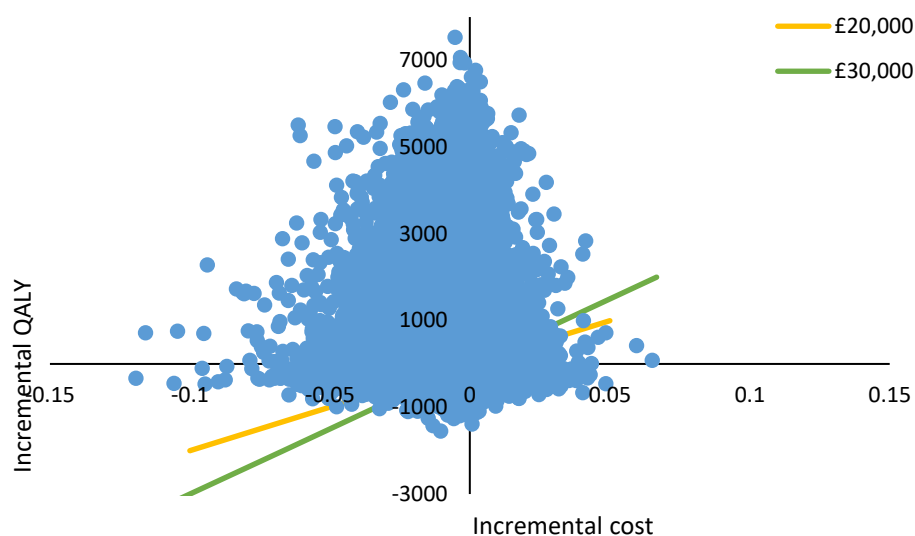
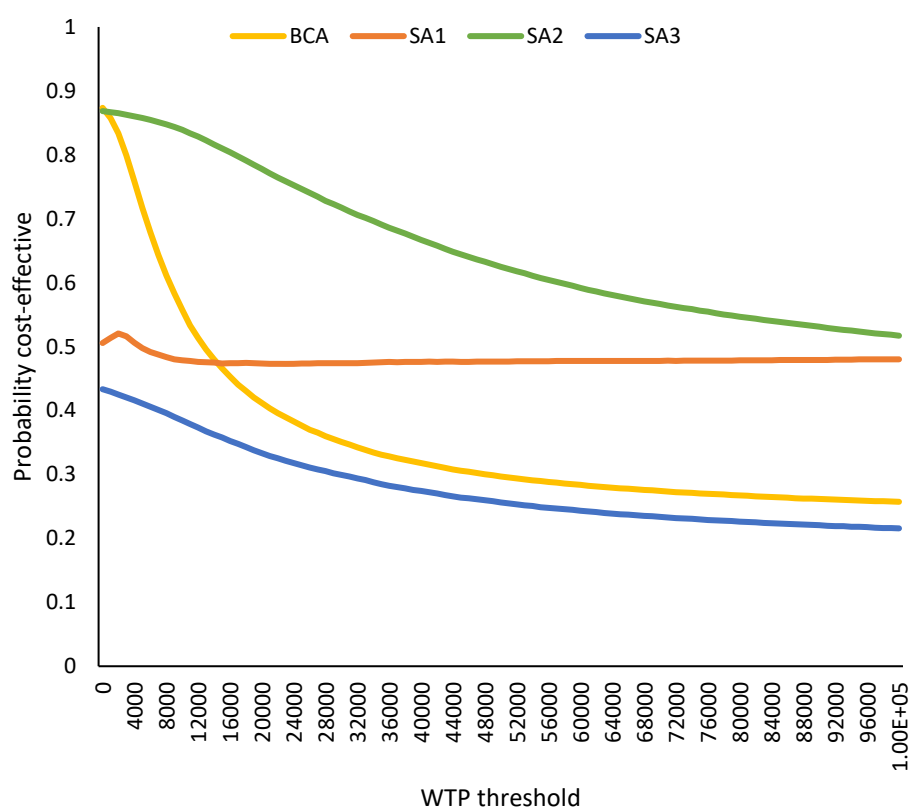


Figure 7-7 CEAC for all scenarios



The CEACs in figure 7-7 summarise the impact of uncertainty in the model parameters on the cost-effectiveness of digitisation. Rather than present a single INB value CEACs allow decision

makers to understand the likelihood of cost-effectiveness and so make better-informed decisions about the adoption of digitisation in pathology. Figure 7-7 shows the SA2 scenario, where costs and QALYs are based on 12months treatment, to consistently produce probabilities of cost-effectiveness that are superior to those for the other scenarios across all WTP thresholds. The SA3 scenario where model probability estimates are based on a single head-to-head study has a consistently lower probability of cost-effectiveness across all the WTP thresholds. NICE recommend a WTP of £20,000 to £30,000 per QALY, decision makers in the UK will in particular want to know the probability of cost-effectiveness at these WTP thresholds. These are summarised in table 7-11.

Table 7-11 Probability of cost-effectiveness at NICE WTP thresholds

	WTP threshold	
	£20,000	£30,000
BCA	0.412	0.351
SA1	0.473	0.474
SA2	0.778	0.717
SA3	0.334	0.299

Out of the four scenarios, three generate probabilities of cost-effectiveness at both WTP thresholds that are less than 0.5. Under BCA, SA1 and SA3 there is a greater likelihood of digitisation not being cost-effective. Under SA2, there is a greater than 70% chance of digitisation being cost-effective.

7.4.3 Value of Information Analysis

The VOI analysis provides evidence to support further research and identifies where this should be focused. Patient level EVPI for all scenarios and EVPPI for the base case and single study scenarios were carried out.

7.4.3.1 EVPI results

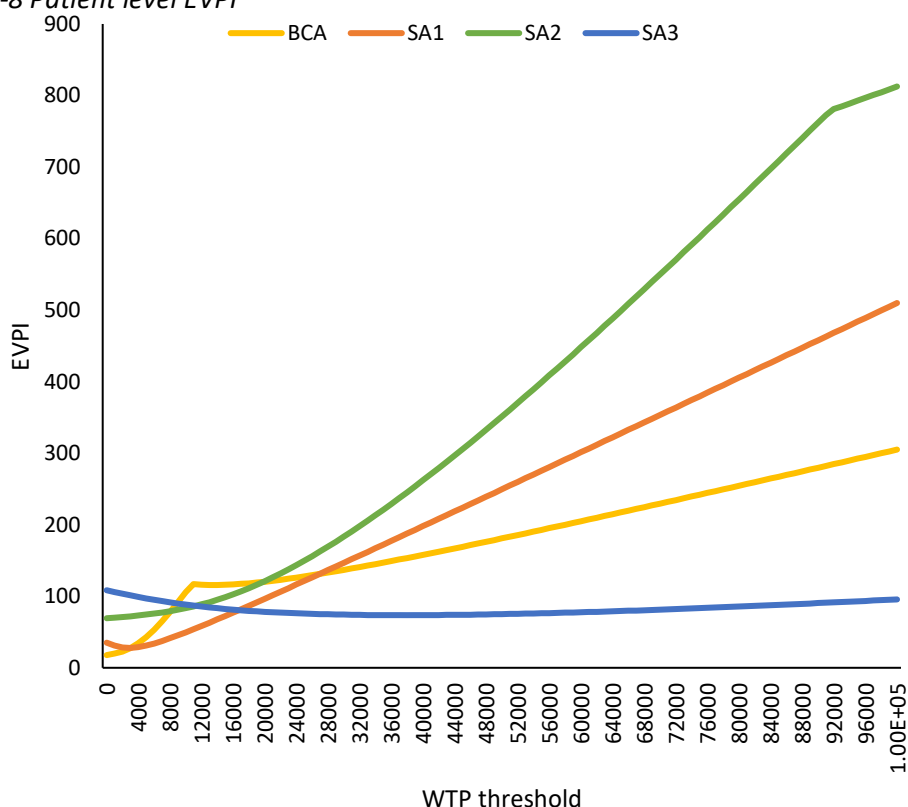
A graphical representation of the patient EVPI at a range of WTP thresholds is given in figure 7-8. The value of further research is generally highest for the SA2 scenario where 12months of Herceptin treatment is given and lowest for the SA3 scenario where single study estimates of model probabilities are used. Table 7-12 summarises the EVPI values at both the NICE WTP thresholds for each scenario.

At the £20,000 WTP threshold EVPI for the base case analysis is very close to that for SA2, however the gap widens as the threshold increases. The SA3 scenario remains associated with the lowest EVPI values.

Table 7-12 EVPI at NICE WTP thresholds

	WTP threshold	
	£20,000	£30,000
BCA	£120.36	£136.84
SA1	£96.53	£146.66
SA2	£121.09	£183.33
SA3	£78.26	£74.32

Figure 7-8 Patient level EVPI



To achieve a meaningful value of EVPI that can be compared to the cost of further research, the population EVPI is calculated by multiplying the patient EVPI by the UK incidence rate of invasive breast cancer. The population EVPI was calculated for each scenario and at both thresholds over a three, five and ten year horizon for comparison purposes. CancerResearchUK (2018) determined an annual incidence of 54,751 in 2015 for invasive breast cancer cases in the UK. This was discounted by 3.5%. It was assumed that the HER2 IHC test was carried out on all invasive breast cancer patients, in reality this is not the case due to the presence of age and comorbidity factors. Where clinicians know on outset that a recommendation for Herceptin treatment cannot be made regardless of the patient's HER2

status they do not expose them to the HER2 testing process. The results of the population EVPI are given in table 7-13.

Table 7-13 Population level EVPI

Model	WTP	Time Horizon		
		3 years	5 years	10 years
BCA	£20,000	£25,052,883	£36,344,333	£61,396,662
	£30,000	£28,481,329	£41,317,996	£69,798,695
SA1	£20,000	£20,093,080	£29,149,123	£49,241,759
	£30,000	£30,526,064	£44,284,302	£74,809,691
SA2	£20,000	£25,205,117	£36,565,180	£61,769,740
	£30,000	£38,159,158	£55,357,667	£93,515,981
SA3	£20,000	£16,290,172	£23,632,228.31	£39,922,040
	£30,000	£15,469,277	£22,441,352.21	£37,910,287

7.4.3.2 EVPPI

The value of eliminating uncertainty around the model parameters in both the base case and SA3 scenarios is investigated with EVPPI. This identifies where further research should be focused. EVPPI was investigated for five groups of parameters, the first valued the elimination of uncertainty in all the p1 (CM) probabilities, the second valued the elimination of uncertainty around test performance and the third and fourth investigated elimination of uncertainty around the QALY values for the TP and FN cases. The results of this analysis are given in table 7-14.

Table 7-14 EVPPI results

Model parameters	BCA		SA3	
	WTP threshold		WTP threshold	
	£20,000	£30,000	£20,000	£30,000
CM probabilities (p1 _x)	£28.47	£10.91	£0.00	£0.00
Test performance (LogOR _x)	£3.08	£1.48	£65.30	£49.22
TP QALYs	£68.56	£73.49	£0.00	£0.00
FN QALYs	£48.65	£50.22	£0.00	£0.00

The greatest value in research under the SA3 scenario is around test

performance. The SA3 scenario measures these uses a single study, contrarily the EVPPI for test performance under the base case analysis were log(OR_x) are based on meta-analysis is considerably lower. The base case EVPPI suggests the greatest value in research is around the QALY estimates that are used to populate the model.

7.5 Discussion

The aim of the economic evaluation was to explore the effects of pathology lab digitisation together with image analysis algorithms on the clinical and cost outcomes for invasive breast cancer patients when scoring HER2 IHC. A decision modelling approach to the economic evaluation was taken and a summary of the cost effectiveness probabilities is given in table 7-11. The majority of results favour CM and do not support the adoption of DP IA. The probability of cost-effectiveness falls below 0.5 at the WTP thresholds for all the scenarios investigated except for SA2, where the costs and QALYs of Herceptin are based on the 12 months treatment plan. This model includes the current recommended treatment duration. The cost of Herceptin under this scenario is much greater than at base case and the range of QALY estimates is lower. In the cost effectiveness plane for SA2 in figure 7-5, many of the estimates of incremental costs and QALYs fall in the bottom two quadrants where DP IA would be considered a cost saving intervention. Since the cost of Herceptin is higher under this scenario, the total cost saving across the model is much greater than that seen at base case. The cost saving is more likely to balance out the monetary value of the QALYs lost so that INB is greater than 0 and DP IA is more likely to be considered cost-effective.

Due to the ethical implications it can be challenging to make cost-effectiveness decisions based on results that fall in the south-west quadrant, in these cases the intervention's cost-effectiveness is associated with a reduction in clinical effectiveness. DP IA will reduce the QALYs gained for some patients that would have otherwise gained under the CM scenario (Dowie et al., 2015). Discussion has taken place around this issue with methods suggested to determine the cost-effectiveness of interventions that fall in the south-west quadrant. One such proposal is the use of a different threshold. The compensation a patient is willing to accept (WTA) for a health loss is greater than the amount they would be willing to pay for the same health gain as demonstrated in the review by O'Brien et al. (2002). Eckermann (2015) suggest the use of what they refer to as a 'kinked threshold' that is based on WTA in the south west quadrant and WTP in the north east quadrant of the cost effectiveness plane. Under such an approach, interventions are less likely to be cost effective if they fall in the south west quadrant. However when taking a societal perspective the aim is to maximise total health benefits to society within a limited budget so that the opportunity cost of investing in interventions that fall in the south west quadrant must be considered (Klok and Postma, 2004).

The increased probability of cost-effectiveness under SA2 compared to the other scenarios is driven by the Herceptin treatment regimen. There is no dispute around the value of Herceptin when given to the correct patients but there is much debate and ongoing research around the recommended length of treatment to reach full benefits with minimum cardiotoxicity. The HERA trial; a three arm randomised control trial compared no Herceptin treatment with Herceptin for 1 year and Herceptin for 2 years (ICR, 2017). Whilst the trial found an improvement in patient outcomes when patients were given 1 year of Herceptin vs no Herceptin; there was no benefit in increasing the duration of Herceptin treatment to two years and evidence indicated a higher possibility of cardio toxic events with 2 years of treatment (Goldhirsch et al., 2013). NICE guidelines recommend that Herceptin is given for 1 year of treatment post chemotherapy (NICE, 2006b) and this is still considered the standard treatment pattern for HER2 positive breast cancers. Research to determine if shorter periods of Herceptin treatment will produce results equivalent to one year are ongoing with investigations into both 6 months and 9 week treatment courses.

There is not yet any full publication on the Persephone trial comparing 6 month to 12 month Herceptin treatment but early indications suggest the non-inferiority of 6 months, reducing both cardio-toxicity and the costs of treatment (Earl et al., 2018). Joensuu et al. (2009)'s trial comparing 9 weeks of adjuvant therapy with or without Herceptin presented evidence supporting the use of Herceptin. Whilst there are no head to head trials comparing 9 weeks to 1 year of Herceptin treatment; Clarke et al. (2017)'s Network Meta-Analysis suggested cost savings and QALY increases when comparing 12 months against 9 weeks of treatment; hence showing the 9 week period to dominate the 12 months. Reducing the length of Herceptin treatment whilst maintaining the outcomes seen with 12 months of treatment will guard patients from cardio-toxic effects. These are only experienced during the treatment period; with risk of these events returning to normal levels once treatment ceases (Onitilo et al., 2014). A shorter period of treatment coupled with accuracy improvements will undoubtedly result in cost-savings to the NHS and allow patients to set on the path to remission sooner.

A perceived limitation of the economic evaluations is the lack of suitable accuracy data to populate the model. The meta-analysis were limited in the number of studies that could be included due to the various criteria that had to be considered. There is undoubtedly a need for further research, particularly around the test performance of the techniques used for scoring. The population EVPI is an upper bound of what would be considered suitable funding for further research. The population EVPI ranged from just 15 to 90 million depending on

the scenario, WTP threshold and time horizon used in the analysis. Certainly, the benefits of additional research will outweigh the cost of further research, as it is expected that the cost of further research will be a lot lower than these values. It is worth noting that all invasive breast cancer female patients will not be exposed to this test and the population EVPI analysis could have taken this into consideration. However based on data collected at UHCW these cases represent a small proportion of breast cancer patients and so is likely to have little impact on the calculated population EVPI. Breast cancer is the second most common cancer in the UK, investment in research that could bring benefits to this large group of patients should be supported.

EVPI supports further research, decisions around where research funding should be focused were also determined. The EVPPI of four sets of model parameters were valued across the SA3 scenario based on a single study and the base case scenario. Interestingly the highest values for EVPPI under the SA3 scenario are around the test performance of DP IA compared to CM. These are largely diminished under the base case scenario, most likely because the test performance is based on synthesised studies. Whilst the number of eligible studies in the systematic review has been repeatedly highlighted as a limitation of the review, the EVPPI is low suggesting no further research is needed. Under the base case scenario the high EVPPI values are associated with QALY estimates for the TP and FN cases, these are taken from Clarke et al. (2017). They used a network meta-analysis to estimate the total QALYs per patient for each of the three arms of their cost-effectiveness analysis, no Herceptin treatment, 9 weeks treatment and 12months treatment. However they do highlight large uncertainties around the QALY values as a limitation of their study, explaining the EVPPI results. Further research on determining QALY values associated with Herceptin treatment are needed. In addition it was particularly challenging to find estimates in the literature of the QALY values for the TN and FP cases in the model.

The evaluation did not consider the costs associated with the use of image analysis algorithms. To an extent the absence of a digital pathology cost deliberation is justifiable as the use of image analysis assumes digital systems are already in place with previous rationalisation and defence. In reality this is not the case; the cost of the introduction of image analysis algorithms for any biomarker or the cost effectiveness of digital pathology as a standalone system have not been studied elsewhere.

This research shows promise and can be used to define the feasibility of transforming pathology labs through digitisation. Further research incorporating the test into a diagnostic

RCT and following patients through from diagnosis through to Herceptin treatment, capturing all costs and clinical outcomes is needed to make an informed judgement on digitisation from the perspective of the HER2 IHC test. Understandably this requires time and resources and so more interim robust head-to-head cohort studies are recommended from which accuracy estimates can be meta-analysed and results used to inform models. These are straightforward to carry out and less resource intense. The results from such a study can be compared to those from a diagnostic RCT when available, strengthening the position in favour or against digitisation.

While the picture is not clear for this biomarker it is important to keep in mind that digitisation will not only be used for image analysis HER2 scoring but in most cases alone producing enhancements in system work flow practices and total throughput in pathology labs. In addition, it will be used alongside scoring algorithms for a vast array of biomarkers, across many diseases including cancers; here there may be potential for a more superior refinement of accuracy. Essentially digital pathology cannot be viewed solely from the HER2 angle but as a system likely to have effects on a wide range of areas in healthcare. The cost-effectiveness of this technology can be proved through many other avenues whether by focusing on digitisation alone or with the parallel use of image analysis algorithms; the research in this area is lacking.

7.6 Conclusion

In conclusion, it is not clear if HER2 IHC scoring using DP IA represents a cost-effective alternative to CM. At the current UK guidelines for treatment and scoring thresholds, DP could possibly be cost-effective however the ethical implications of the QALY loss must be considered when a decision on investment is made. If the length of Herceptin treatment is reduced as the research currently indicates, DP cannot be considered a cost-effective intervention from the perspective of this particular biomarker.

Chapter 8: General discussion and conclusion

8.1 Overview of thesis and main findings

The work presented through this thesis attempted to begin to address the research gaps identified by the systematic review in chapter 2.

The systematic review highlighted the paucity of published research around the economic evaluations of innovation within pathology practice. Whilst this was perhaps not the case for new cancer tests it was particularly clear that no economic analysis had been carried out to justify changes in systems and processes within pathology or to justify the introduction of new technologies designed to maximise efficiency, productivity and enhance the accuracy of pathologists' reporting. Even so the few studies that were eligible for inclusion were mainly related to testing, they were not specifically evaluations of new tests but the majority focused on the adaptation of existing testing strategies to improve accuracy. There were no economic evaluations of generic process pathology lab interventions that have a diffuse effect within and beyond the pathology lab. Based on the review findings in chapter 2, the thesis focused on identifying, measuring and valuing the costs and benefits of digital pathology when used within the breast cancer pathway.

No published work has previously attempted to measure both the costs and benefits of digital pathology to the cancer patient.

8.1.1 The breast cancer pathway

To evaluate the impacts of pathology digitisation for breast cancer patients, an appreciation of the pathology activities, their position within the breast cancer pathway and how they interact with events taking place across other medical specialities that feed into patient outcomes had to first be gained. This was achieved by reviewing the published breast cancer guidelines and interviewing the full spectrum of clinicians involved in breast cancer management and treatment at UHCW. The resultant breast cancer pathway output is described in chapter 4. As there were no discrepancies between the recommended guidelines and clinical practice at UHCW, with information from the latter supplementing what is learnt from the former, the pathway represented what could be interpreted as good practice clinical care within a large teaching hospital. This pathway was used as a foundation for the construction of the DES model in chapter 5, and a section within it was the basis for the second decision model developed in chapter 7.

In reality, the breast cancer pathway illustrated can be used as an information source on which to base future research. Researchers both within and outside of pathology, evaluating

interventions that have diffuse consequences across the breast cancer pathway can use the illustrated diagrams to gain an overview of the chain of events between the intervention they are evaluating and patient outcomes. This helps researchers understand the sections of the overall pathway that are relevant and should be considered in the evaluation to gain a holistic view of the gains or losses for breast cancer patients. This is particularly useful when evaluating system changes or complex interventions that have more than one impact on the pathway and thus several ways by which they can influence patient benefits. In reality the processes described in chapter 4 can be duplicated to depict pathways of good practice clinical care for all cancers or even across all disease areas to provide a readily available resource that can be used as a base in future economic evaluations.

Chapter 3 summarises the published literature on the digital pathology impacts for the breast cancer patient. These varied and included both time savings and accuracy gains. Nonetheless these were in many cases hypothetical discussions based on very limited evidence and a research gap was certainly identified. Other than studies that measured the concordance in reporting for some of the characteristics of breast cancer there were no good quality synthesis or head-to-head studies that explored the time implications of digitisation when compared to standard pathology practice.

Through knowledge of both the impacts of digitisation for breast cancer patients learnt in chapter 3 and information understood about the pathway in chapter 4, it was possible to construct the process chain network of events for pathology digitisation that feeds into patient outcomes. Familiarity of this network of events was vital to constructing the DES model to ensure that pathology always remained at the centre of all modelling construction decisions.

Beyond chapter 4 and the breast cancer pathway illustrated, the remaining chapters in the thesis were split between two economic evaluations of digital pathology. Firstly, chapter 5 focused on the evaluation of digital pathology from a broad perspective as a generic process by taking a complete view of the breast cancer pathway. Chapters 6 to 7 evaluate the cost-effectiveness of digital pathology from a narrower perspective as a targeted process, focusing on a single test used within the breast cancer pathway

8.1.2 Economic evaluation using DES modelling

Digital pathology is a technological intervention within the pathology workflow, introduced to enhance efficiency and productivity as well as contribute to improved accuracy by supporting pathologists to report disease characteristics. The published literature describe

many points of its influence within the breast cancer pathway as summarised in chapter 3. DES modelling allowed the simultaneous evaluation of the processes in the breast cancer pathway influenced by digitisation to gain a holistic view of the impacts of digital pathology for breast cancer patients rather than the impact on one chain of processes. The model was first developed conceptually based on the full breast pathway illustrated in chapter 4. The conceptual model is a summary of the more detailed pathway and is designed to reflect the processes that are influenced by digitisation and the chain of events that contribute to patient outcomes.

Construction of the DES model is in itself an output of the thesis. The DES model has been designed to model pathology and all its interactions within the breast cancer pathway. It can be used as a tool for future evaluations of any pathological intervention introduced within the breast cancer pathway. One of the main limitations of the economic evaluation in chapter 5 is the lack of primary research investigating the impacts of digitisation so that the economic evaluation is based on a series of hypothetical scenarios rather than good quality research evidence. As research is published and becomes available the model can be updated to reflect the most recent findings in order to validate or refute the preliminary findings presented in chapter 5. Furthermore, in the same way as for digitisation, evaluations of other generic pathology processes that may influence the breast cancer pathway can be carried out using the DES model. The DES model is too complex and unnecessary for the evaluation of targeted pathology processes e.g. modifications in testing as stand-alone interventions, the cost-effectiveness of these can be explored independently using simpler modelling approaches as was done for the HER2 test in chapters 6 and 7. However the costs and effectiveness data related to targeted processes can be explored within the network of processes of an upper level generic pathology process.

The motivation for the economic evaluation carried out in chapter 5 was largely exploratory for this reason a cost-consequence analysis was chosen. The evaluation starts to build a picture of the combined impacts of pathology digitisation on the breast cancer pathway. Several scenarios for digital pathology were explored, they reflected a series of hypothetical impacts of pathology digitisation. They were all equal on changes observed in the test results reported for HER2 and grade but differed on the time savings they predicted. The first explored an overall 10% time saving across all pathology activities, the second explored consequences if 50% of CBs are reported on in 1 day and the third scenario, if 100% are reported on in 6 hours.

Evidence of time savings in the pathway were observed for all breast cancer patients under each of the DP scenarios when compared to CM. Time savings undoubtedly contribute to achieving government cancer waiting time targets however it was difficult to map the time savings in the pathway to the specific government targets. Cancer waiting time targets start from referral in primary care, this phase of the patients' care pathway is not incorporated into the DES model as it is not relevant to evaluations of pathology. Nonetheless there is evidence of an overall time saving in the breast pathway due to the combined impact of the various effects of pathology digitisation. However this needs to be explored further by using model parameters and estimations of time savings that are based on good quality research rather than clinician estimations as is currently the case.

MDT outputs of the DES model in chapter 5 showed evidence of a decrease in the number of MDT discussions per breast cancer case across all of the DP scenarios when compared to CM. This may not in particular be caused by enhanced efficiency within the MDT meeting itself but a consequence of enhanced efficiency within the pathology lab, demonstrating how pathology modernisation can have positive benefits beyond the lab. Test results are available sooner and breast cases can be discussed in a single MDT meeting rather than going back on forth between MDT and the pathology lab due to test results not being available. Even though it was beyond the scope of the model to be measured, fewer MDT discussions per case, frees up clinician time for the discussion of more complex cases in the meeting or even could possibly shorten the overall length of the meeting itself, freeing up clinicians time for other activities. Modernisation in pathology can have far reaching impacts beyond those outlined in the process chain networks in figure 4-15.

The model also extrapolated the impacts of pathology digitisation on test results accuracy to the treatment packages prescribed. The model was able to measure the joint impact of changes in accuracy for both HER2 scoring and the grade determined since these are both considered when recommending chemotherapy. There was an evident decrease in the number of patients prescribed combined chemotherapy and Herceptin and an increase in the number of patients prescribed chemotherapy alone. These are preliminary findings and further exploration of the time outputs are needed as primary research becomes available that compares CM and DP accuracy.

Across all three DP scenarios there was evidence of cost savings with pathology digitisation when compared to the use of conventional microscopes. Modifications in treatment packages, reductions in the number of MDT discussions per case and a decrease in the

number of HER2 FISH tests requested, all contribute to the cost savings observed. Modernising pathology labs through digitisation is evidence of a cost-saving intervention.

A limitation of the DES model is that it didn't go as far as mapping changes in clinical processes such as the treatment prescribed to patient outcomes or value time savings to the patient. However it presented an example of how DES modelling can be used in the evaluation of a systems intervention such as pathology digitisation. The model is available and can be used as tool in future evaluations of pathology interventions within the breast cancer pathway.

8.1.3 HER2 IHC accuracy with digital pathology

The dearth of primary research that compares CM to DP was highlighted in chapter 5, this motivated the comparison of HER2 IHC accuracy that was carried out in chapter 6. Accuracy estimates of sensitivity and specificity for each of CM and DP when scoring HER2 IHC were measured using a bivariate meta-analysis and compared against each other in ROC space.

A strength of the systematic review that was carried out is the inclusion of head-head studies only that compare the two tests directly in the same sample reducing the risk of bias due to confounding. However this limited the number of studies that were eligible for inclusion in the review especially when studies were grouped according to the underlying thresholds used for scoring in each study.

Using bivariate meta-analyses techniques rather than univariate meta-analysis was a strength of this review. Bivariate meta-analyses preserve the underlying relationship between sensitivity and specificity when pooling results and provide better estimations of these accuracy estimates than the use of univariate meta-analyses. A challenge within this study was the presence of the equivocal score within the HER2 category, this group was simply excluded from 2x2 tables when estimating sensitivity and specificity and pooled separately. This is a limitation of the analysis, since the relationship between the probability of an equivocal score and sensitivity or specificity was not considered.

The evidence generated by the meta-analyses did not necessarily support the superiority of digitisation over conventional microscope approaches. Using the 2013 threshold data it is possible to conclude that digitisation can potentially signify a more precise approach to scoring HER2 IHC than conventional microscopes, first through the reduction of equivocal scores and secondly by reducing the confidence region around the mean. However it cannot be established that there are improvements in accuracy of IHC scoring with digitisation when

compared to conventional microscopes. For the 2008 threshold data there was no difference in sensitivity and specificity between CM and DP.

The value of IA for scoring HER2 IHC can be disputed based on the meta-analysis results. The evidence does not support the hypothesis that DP IA results in superior accuracy when scoring HER2 IHC. All the studies that were included in the systematic review specifically had to include IA as part of their estimation of HER2 scores in the DP arm. Further analysis could include studies that used DP alone without IA to support scoring. In this situation DP is identical to CM except the pathologist is viewing computer slides rather than glass slides.

The intention is not to dismiss IA, but its value for this particular biomarker is questionable. However IA algorithms can be generated for all sorts of biomarkers beyond breast cancer. The value of IA could be greater when being used to support the pathologist in scoring a disease characteristic such as grade where there is known to be wide inter- and intra- observer variability. Meta-analysis can only be carried out when the primary research becomes available, it was evident, for breast cancer at least that this information is lacking.

Sensitivity and specificity outputs of the bivariate meta-analysis model that were based on the 2008 and 2013 thresholds were compared. The 2008 threshold data showed sensitivity to be lower than that at the 2013 threshold but specificity to be higher under the 2008 threshold demonstrating the ability of the 2013 thresholds at better identifying positive HER2 cases. This supports the continual research around cut-off thresholds to improve accuracy. Since this study the ASCO/CAP guidelines have once more been updated however it would not have been possible to estimate test performance based on these since undoubtedly the published research is still very early in development.

8.1.4 Economic evaluation of digital pathology for HER2 scoring

Measuring changes in accuracy are of little value if not measured directly to the patient. In this economic evaluation the targeted pathology process i.e. HER2 IHC test was mapped onto the clinical process i.e. Herceptin treatment to measure the outcomes of digitisation to the patient. Test performance was used to inform a decision tree model comparing the costs and benefits of digital pathology image analysis to the use of conventional microscopes when scoring HER2 IHC. DP IA for scoring HER2 IHC compared to CM is not a cost-effective alternative, this is largely due to the QALY losses experienced with DP. These results support those generated in the previous chapter, DP IA greatest influence is not likely to be on the HER2 biomarker.

The study confirms that the benefit of digitisation will come from efficiency and productivity improvements and not necessarily from accuracy. However these results need only be interpreted from the breast cancer perspective. In reality different results may be observed when considering other cancers or disease areas that digitisation of pathology is expected to influence.

8.2 Strengths and limitations of research

There are several strengths to the research undertaken. Firstly, the use of DES modelling for the evaluation of complex health interventions was demonstrated and the model can be used for future research. The use of bivariate rather than univariate meta-analysis for estimations of sensitivity and specificity, however a limitation was the omission of equivocal scoring group from these analyses. A third strength is the level of detail that was gathered around the cancer pathway through guidelines and clinician interviews so that the models developed accurately represent the breast cancer pathway.

The main limitation of the research and this is particularly the case for the DES model is the hypothetical nature of the scenarios that were drawn up. Whilst the evaluations suggest DP is cost and time savings, these are hypothetical conclusions only and are not robust enough on which to base investment decisions.

8.3 Recommendations for policy

Digitisation has been at the centre of plans to transform the NHS since 2013, digital pathology is part of this wider plan of modernisation. Pathology departments have repeatedly cited access to funding and high upfront costs as a barrier to the introduction of digitisation (Williams et al., 2018b). When evaluating the impact of DP for HER2 accuracy there was evidence of cost savings (albeit QALY losses) and costs savings were also identified from the evaluation of DP as a whole. These are just cost savings from the breast cancer perspective, if similar evidence is presented for other cancers, the cost saving implications of DP could be immense. Based on the UHCW business case the cost of digital pathology per slide is less than £0.50, if all disease areas that are impacted by digitisation in pathology are considered, there will certainly be evidence that the cost savings as a result of digitisation will outweigh the cost of investment.

There is certainly a case for the introduction of digital pathology systems based on this impact alone, since in reality DP without IA is expected to at least be equivalent to CM, as they are both simply different mediums for viewing the same slide. It is IA that will really make the difference to accuracy and impact patient outcomes.

8.4 Recommendations for research

Primary research data around the effectiveness of digital pathology for each of the processes within the breast cancer pathway it is expected to influence was very limited. This research is preliminary to any undertaking that evaluates digital pathology as a whole system intervention to gain a full picture of its implications. While accuracy is important efficiency is overlooked in many cases in the published literature. Further research should focus on exploring time implications and how these translate into patient benefits since the study undertaken suggests that the greatest value of digitisation will be through time savings. The value of reducing time to diagnosis for cancer patients has not previously been measured and this could be considered in future research. Time savings contribute to early cancer diagnosis and are expected to improve survival outcomes, however these were not measured in the study and further research is needed to confirm that the few days saved in the pathway will in fact have this consequence which is the motivation behind government cancer waiting time targets.

DES modelling has been used for the evaluation of pathology services within a cancer pathway. DES modelling can be considered for the evaluation of activities within the pathology lab to gain a granular understanding of the impacts of DP on workflow and the challenge of understaffing.

Whilst the focus of this thesis has been on the effects of digitisation to patients, there are benefits beyond those to pathology staff. Since workforce constraints have been highlighted as a major challenge for the future of pathology, future research could focus on the implications of digitisation in compensating for staff shortages and overcoming the burden of increased demand for services.

8.5 Conclusions

The thesis presented an example of an economic evaluation of a generic process whose impacts crosses multiple disease pathways. Innovation in pathology is expected to have positive benefits for cancer patients. The preliminary research undertaken has shown evidence of time savings for breast cancer patients and cost savings to the NHS as a direct result of pathology digitisation.

However when evaluating the impact of digitisation specifically on the accuracy of the HER2 biomarker, the cost-effectiveness analysis did not support digital pathology adoption.

Further primary research is needed to measure and value the impact of pathology digitisation on time in the pathway and the accuracy of other cancer biomarkers.

References

- ABDEL-RAHMAN, M., STOCKTON, D., RACHET, B., HAKULINEN, T. & COLEMAN, M. 2009. What if cancer survival in Britain were the same as in Europe: how many deaths are avoidable? *British journal of cancer*, 101, S115-S124.
- ABS 2009. Surgical guidelines for the management of breast cancer. 35, S1-S22.
- AEFFNER, F., WILSON, K., MARTIN, N. T., BLACK, J. C., HENDRIKS, C. L. L., BOLON, B., RUDMANN, D. G., GIANANI, R., KOEGLER, S. R., KRUEGER, J. J. A. O. P. & MEDICINE, L. 2017. The gold standard paradox in digital image analysis: manual versus automated scoring as ground truth. 141, 1267-1275.
- AL-JANABI, S., HUISMAN, A., WILLEMS, S. & VAN DIEST, P. J. H. P. 2012. Digital slide images for primary diagnostics in breast pathology: a feasibility study. 43, 2318-2325.
- AL-JANABI, S., HUISMAN, A. & VAN DIEST, P. J. 2012. Digital pathology: current status and future perspectives. *Histopathology*, 61, 1-9.
- ALL-PARTY PARLIAMENTARY GROUP. 2018. *A mixed picture: An inquiry into geographical inequalities and breast cancer* [Online]. All-Party Parliamentary Group on Breast Cancer. Available: file:///marlon.ads.warwick.ac.uk/User55/u/u1111322/Desktop/appgbc_a_mixed_picture.pdf [Accessed 12/10/2018].
- ALLGAR, V. & NEAL, R. 2005. Delays in the diagnosis of six cancers: analysis of data from the National Survey of NHS Patients: Cancer. *British journal of cancer*, 92, 1959-1970.
- AMES, V. & BRITTON, P. D. 2011. Stereotactically guided breast biopsy: a review. *Insights into imaging*, 2, 171-176.
- ANDREANO, A., REBORA, P., VALSECCHI, M. G., RUSSO, A. G. J. B. C. R. & TREATMENT 2017. Adherence to guidelines and breast cancer patients survival: a population-based cohort study analyzed with a causal inference approach. 164, 119-131.
- ANDREU, F. J., SÁEZ, A., SENTÍS, M., REY, M., FERNÁNDEZ, S., DINARÈS, C., TORTAJADA, L., GANAU, S. & PALOMAR, G. 2007. Breast core biopsy reporting categories—An internal validation in a series of 3054 consecutive lesions. *The Breast*, 16, 94-101.
- ARE, C., CANIGLIA, A., MOHAMMED, M., SMITH, L., CUMMINGS, C., LECOQ, C., BERMAN, R., AUDISIO, R. & WYLD, L. 2018. Global variations in the level of cancer-related research activity and correlation to cancer-specific mortality: Proposal for a global curriculum. *Annals of Surgical Oncology*, 25, 594-603.
- BAHREINI, F., SOLTANIAN, A. R. & MEHDIPOUR, P. 2015. A meta-analysis on concordance between immunohistochemistry (IHC) and fluorescence in situ hybridization (FISH) to detect HER2 gene overexpression in breast cancer. *Breast cancer*, 22, 615-625.
- BARTON, P., BRYAN, S., ROBINSON, S. J. J. O. H. S. R. & POLICY 2004. Modelling in the economic evaluation of health care: selecting the appropriate approach. 9, 110-118.
- BEASTALL, G. H. J. T. C. B. R. 2008. The modernisation of pathology and laboratory medicine in the UK: networking into the future. 29, 3.
- BEJNORDI, B. E., VETA, M., VAN DIEST, P. J., VAN GINNEKEN, B., KARSSMEIJER, N., LITJENS, G., VAN DER LAAK, J. A., HERMSEN, M., MANSON, Q. F. & BALKENHOL, M. J. J. 2017. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. 318, 2199-2210.
- BILOUS, M. 2010a. Breast core needle biopsy: issues and controversies. *Modern Pathology*, 23, S36-S45.
- BILOUS, M. 2010b. Breast core needle biopsy: issues and controversies. *Modern Pathology*, 23, S36.
- BISHOP, J. W., MARCELPOIL, R. & SCHMID, J. 2002. Machine scoring of her2/neu immunohistochemical stains. *Analytical and Quantitative Cytology and Histology*, 24, 257-262.

- BLANK, P. R., SCHWENKLENKS, M., MOCH, H. & SZUCS, T. D. 2010. Human epidermal growth factor receptor 2 expression in early breast cancer patients: a Swiss cost-effectiveness analysis of different predictive assay strategies. *Breast Cancer Research and Treatment*, 124, 497-507.
- BLOOM, K. & HARRINGTON, D. 2004. Enhanced accuracy and reliability of HER-2/neu immunohistochemical scoring using digital microscopy. *American Journal of Clinical Pathology*, 121, 620-630.
- BOIESEN, P., BENDAHL, P.-O., ANAGNOSTAKI, L., DOMANSKI, H., HOLM, E., IDVALL, I., JOHANSSON, S., LJUNGBERG, O. & RINGBERG, A. J. A. O. 2000. Histologic grading in breast cancer: reproducibility between seven pathologic departments. 39, 41-45.
- BORRELLI, C., COHEN, S. & DUNCAN, A. 2016. NHS breast screening programme: clinical guidance for breast cancer screening assessment.
- BREAST CANCER CARE. 2013. *Understanding your pathology results* [Online]. Breast Cancer Care. Available: https://www.breastcancercare.org.uk/sites/default/files/publications/pdf/bcc161_understanding_your_pathology_results_web.pdf [Accessed 27/09/2017].
- BREAST CANCER CARE. 2016. *Primary Breast Cancer* [Online]. Breast Cancer Care. Available: https://www.breastcancercare.org.uk/sites/default/files/publications/pdf/primary_resource_pack_2016_web.pdf [Accessed 27/09/2017].
- BREAST CANCER CARE. 2018. *Cancer grade and size* [Online]. Available: <https://www.breastcancercare.org.uk/information-support/facing-breast-cancer/diagnosed-breast-cancer/cancer-grade-size> [Accessed 06/10/2018].
- BRENNAN, M. E., TURNER, R. M., CIATTO, S., MARINOVICH, M. L., FRENCH, J. R., MACASKILL, P. & HOUSSAMI, N. 2011. Ductal carcinoma in situ at core-needle biopsy: meta-analysis of underestimation and predictors of invasive breast cancer. *Radiology*, 260, 119-128.
- CABANA, M. D., RAND, C. S., POWE, N. R., WU, A. W., WILSON, M. H., ABBODD, P.-A. C. & RUBIN, H. R. J. J. 1999. Why don't physicians follow clinical practice guidelines?: A framework for improvement. 282, 1458-1465.
- CAMERON, D., PICCART-GEHART, M. J., GELBER, R. D., PROCTER, M., GOLDBIRSH, A., DE AZAMBUJA, E., CASTRO, G., UNTCH, M., SMITH, I. & GIANNI, L. 2017. 11 years' follow-up of trastuzumab after adjuvant chemotherapy in HER2-positive early breast cancer: final analysis of the HERceptin Adjuvant (HERA) trial. *The Lancet*, 389, 1195-1205.
- CAMPBELL, W. S., HINRICHS, S. H., LELE, S. M., BAKER, J. J., LAZENBY, A. J., TALMON, G. A., SMITH, L. M. & WEST, W. W. J. H. P. 2014. Whole slide imaging diagnostic concordance with light microscopy for breast needle biopsies. 45, 1713-1721.
- CANCER RESEARCH UK. 2015. *Breast cancer statistics* [Online]. Available: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer#heading-Zero> [Accessed 28/01/2019].
- CANCER RESEARCH UK. 2016a. *Cancer mortality statistics* [Online]. Available: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/mortality#heading-Zero> [Accessed 04/01/2019].
- CANCER RESEARCH UK. 2016b. *Testing Times to Come: An Evaluation of Pathology Capacity Across the UK* [Online]. Available: https://www.cancerresearchuk.org/sites/default/files/testing_times_to_come_nov_16_cruk.pdf [Accessed 16/11/2017].
- CANCER RESEARCH UK. 2017. *Breast Cancer* [Online]. Available: <https://www.cancerresearchuk.org/about-cancer/breast-cancer> [Accessed 27/09/2017].

- CANCER RESEARCH UK. 2018. *Cancer incidence for common cancers* [Online]. Available: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/incidence/common-cancers-compared#heading-Zero> [Accessed 05/01/2019].
- CANCERRESEARCHUK. 2018. *Breast cancer incidence (invasive) statistics* [Online]. Available: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/incidence-invasive#heading-Zero> [Accessed 20/09/2018].
- CAR, L. T., PAPACHRISTOU, N., URCH, C., MAJEED, A., EL-KHATIB, M., AYLIN, P., ATUN, R., CAR, J. & VINCENT, C. 2016. Preventing delayed diagnosis of cancer: clinicians' views on main problems and solutions. *Journal of global health*, 6.
- CARLSON, B. 2008. HER2 TESTS: How Do We Choose? *Biotechnology healthcare*, 5, 23.
- CARLSON, J. J., ROTH, J. A. J. B. C. R. & TREATMENT 2013. The impact of the Oncotype Dx breast cancer assay in clinical practice: a systematic review and meta-analysis. 141, 13-22.
- CARNEIRO, S., FRY, A., IRONMONGER, L., CONNOR, K., ORMISTON-SMITH, N., E HENSON, K., JOHNSON, S., MCPHAILL, S. & ELLISS-BROOKES, L. Variation in the Routes to Cancer Diagnosis and Stage for Ten Cancer Sites. Cancer data and outcomes conference, 2016 Manchester.
- CARSON, I., NICOL, D. M., NELSON, B. L. & BANKS, J. 2005. Discrete-event system simulation. Pearson.
- CARTER, A. J. & NGUYEN, C. N. 2012. A comparison of cancer burden and research spending reveals discrepancies in the distribution of research funding. *BMC public health*, 12, 526.
- CHEN, X., YUAN, Y., GU, Z. & SHEN, K. 2012. Accuracy of estrogen receptor, progesterone receptor, and HER2 status between core needle and open excision biopsy in breast cancer: a meta-analysis. *Breast cancer research and treatment*, 134, 957-967.
- CHILD, M. & GUPTA, L. 2009. A career in pathology. *Student BMJ*, 17.
- CHU, H., NIE, L., COLE, S. R. & POOLE, C. 2009. Meta-analysis of diagnostic accuracy studies accounting for disease prevalence: alternative parameterizations and model selection. *Statistics in medicine*, 28, 2384-2399.
- CIAMPA, A., XU, B., AYATA, G., BAIYEE, D., WALLACE, J., WERTHEIMER, M., EDMISTON, K. & KHAN, A. 2006. HER-2 status in breast cancer: correlation of gene amplification by FISH with immunohistochemistry expression using advanced cellular imaging system. *Applied Immunohistochemistry & Molecular Morphology*, 14, 132-137.
- CLARKE, C. S., HUNTER, R. M., SHEMILT, I. & SERRA-SASTRE, V. 2017. Multi-arm Cost-Effectiveness Analysis (CEA) comparing different durations of adjuvant trastuzumab in early breast cancer, from the English NHS payer perspective. *PLoS One*, 12, e0172731.
- CLARKE, G., EIDT, S., SUN, L., MAWDSLEY, G., ZUBOVITS, J. & YAFFE, M. 2007. Whole-specimen histopathology: a method to produce whole-mount breast serial sections for 3-D digital histopathology imaging. *Histopathology*, 50, 232-242.
- CLAXTON, K. P. & SCULPHER, M. J. J. P. 2006. Using value of information analysis to prioritise health research. 24, 1055-1068.
- COBLENTZ, T. R., MILLS, S. E. & THEODORESCU, D. 2001. Impact of second opinion pathology in the definitive management of patients with bladder carcinoma. *Cancer*, 91, 1284-90.
- COLEMAN, M., GATTA, G., VERDECCHIA, A., ESTEVE, J., SANT, M., STORM, H., ALLEMANI, C., CICCOTALLO, L., SANTAQUILANI, M. & BERRINO, F. 2003. EUROCARE-3 summary: cancer survival in Europe at the end of the 20th century. *Annals of oncology*, 14, v128-v149.

- COOK, D. A. & SKINNER, J. M. 2005. How to perform credible verification, validation, and accreditation for modeling and simulation. *The Journal of Defense Software Engineering*, 18, 20-24.
- COOPER, K. L., MENG, Y., HARNAN, S., WARD, S. E., FITZGERALD, P., PAPAIOANNOU, D., WYLD, L., INGRAM, C., WILKINSON, I. D. & LORENZ, E. 2011. Positron emission tomography (PET) and magnetic resonance imaging (MRI) for the assessment of axillary lymph node metastases in early breast cancer: systematic review and economic evaluation. *Health Technology Assessment (Winchester, England)*, 15, iii.
- CRD 2009. CRD's Guidance for Undertaking Reviews in Health Care.
- CRUZ-ROA, A., GILMORE, H., BASAVANHALLY, A., FELDMAN, M., GANESAN, S., SHIH, N. N., TOMASZEWSKI, J., GONZÁLEZ, F. A. & MADABHUSHI, A. J. S. R. 2017. Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent. 7, 46450.
- CUADROS, M. & VILLEGAS, R. 2009. Systematic review of HER2 breast cancer testing. *Applied Immunohistochemistry & Molecular Morphology*, 17, 1-7.
- D'ANDREA, E., MARZUILLO, C., PELONE, F., DE VITO, C. & VILLARI, P. 2015. Genetic testing and economic evaluations: a systematic review of the literature. *Epidemiologia e Prevenzione*, 39, S45-S50.
- DAVIDSON, M. R., GAZDAR, A. F. & CLARKE, B. E. 2013. The pivotal role of pathology in the management of lung cancer. *JOURNAL OF THORACIC DISEASE*, 5, S463-S478.
- DE ANGELIS, R., SANT, M., COLEMAN, M. P., FRANCISCI, S., BAILI, P., PIERANNUNZIO, D., TRAMA, A., VISSER, O., BRENNER, H. & ARDANAZ, E. 2014. Cancer survival in Europe 1999–2007 by country and age: results of EUROCARE-5—a population-based study. *The lancet oncology*, 15, 23-34.
- DENDUKURI, N., KHETANI, K., MCISAAC, M. & BROPHY, J. 2007. Testing for HER2-positive breast cancer: a systematic review and cost-effectiveness analysis. *Canadian Medical Association Journal*, 176, 1429-1434.
- DEPARTMENT OF HEALTH 2000. *NHS cancer plan: A plan for investment, a plan for reform*, DoH London.
- DEPARTMENT OF HEALTH 2004. *Modernising Pathology Services*. London: Department of Health, 7.
- DEPARTMENT OF HEALTH 2007. *Cancer reform strategy*, Department of Health.
- DEPARTMENT OF HEALTH 2011. *Improving outcomes: a strategy for cancer*.
- DEPARTMENT OF HEALTH 2014. *Living Well For Longer: National Support for Local Action to Reduce Premature Avoidable Mortality*. UK Government London.
- DOBSON, L., CONWAY, C., HANLEY, A., JOHNSON, A., COSTELLO, S., O'GRADY, A., CONNOLLY, Y., MAGEE, H., O'SHEA, D. & JEFFERS, M. 2010. Image analysis as an adjunct to manual HER-2 immunohistochemical review: a diagnostic tool to standardize interpretation. *Histopathology*, 57, 27-38.
- DOEBLER, P. & HOLLING, H. 2015. Meta-analysis of diagnostic accuracy with mada. *R Packag*, 1-15.
- DOWIE, J., KALTOFT, M. K., NIELSEN, J. B. & SALKELD, G. 2015. Caveat emptor NICE: biased use of cost-effectiveness is inefficient and inequitable. *F1000Research*, 4.
- DRUCKER, E. & KRAPPENBAUER, K. J. E. J. 2013. Pitfalls and limitations in translation from biomarker discovery to clinical utility in predictive and personalised medicine. 4, 7.
- EARL, H. M., HILLER, L., VALLIER, A.-L., LOI, S., HOWE, D., HIGGINS, H. B., MCADAM, K., HUGHES-DAVIES, L., HARNETT, A. N. & AH-SEE, M.-L. 2018. PERSEPHONE: 6 versus 12 months (m) of adjuvant trastuzumab in patients (pts) with HER2 positive (+) early breast cancer (EBC): Randomised phase 3 non-inferiority trial with definitive 4-year (yr) disease-free survival (DFS) results. American Society of Clinical Oncology.

- EARLY BREAST CANCER TRIALISTS' COLLABORATIVE GROUP 1998. Polychemotherapy for early breast cancer: an overview of the randomised trials. *The Lancet*, 352, 930-942.
- ECKERMAN, S. 2015. Kinky thresholds revisited: opportunity costs differ in the NE and SW quadrants. *Applied health economics and health policy*, 13, 7-13.
- ELKIN, E. B., WEINSTEIN, K. C., WINER, E. P., KUNTZ, K. M., SCHNITT, S. J. & WEEKS, J. C. 2004. HER-2 testing and trastuzumab therapy for metastatic breast cancer: a cost-effectiveness analysis. *Journal of Clinical Oncology*, 22, 854-863.
- ELLIS, C., DYSON, M., STEPHENSON, T. & MALTBY, E. 2005. HER2 amplification status in breast cancer: a comparison between immunohistochemical staining and fluorescence in situ hybridisation using manual and automated quantitative image analysis scoring techniques. *Journal of clinical pathology*, 58, 710-714.
- ELLIS, I., AL-SAM, S., ANDERSON, N., CARDER, P., DEB, R., GIRLING, A., HALES, S., HANBY, A., IBRAHIM, M. & LEE, A. 2016. Pathology reporting of breast disease in surgical excision specimens incorporating the dataset for histological reporting of breast cancer June 2016.
- ELLISS-BROOKES, L., MCPHAIL, S., IVES, A., GREENSLADE, M., SHELTON, J., HIOM, S. & RICHARDS, M. 2012. Routes to diagnosis for cancer - determining the patient journey using multiple routine data sets. *British Journal of Cancer*, 107, 1220-1226.
- ELMORE, J. G., LONGTON, G. M., PEPE, M. S., CARNEY, P. A., NELSON, H. D., ALLISON, K. H., GELLER, B. M., ONEGA, T., TOSTESON, A. N. & MERCAN, E. J. J. O. P. I. 2017. A randomized study comparing digital imaging to traditional glass slide microscopy for breast biopsy and cancer diagnosis. 8.
- ELSTON, C. W. & ELLIS, I. O. 1991. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology*, 19, 403-410.
- EMMADI, R. & WILEY, E. L. J. I. J. O. S. O. 2012. Evaluation of resection margins in breast conservation therapy: the pathology perspective—past, present, and future. 2012.
- ERNSTER, V. L. & BARCLAY, J. 1997. Increases in ductal carcinoma in situ (DCIS) of the breast in relation to mammography: a dilemma. *JNCI Monographs*, 1997, 151-156.
- EUNETHTA 2014. Meta-analysis of diagnostic test accuracy studies. *European Network for Health Technology Assessment*.
- EUROSTAT. 2018. *Cancer statistics* [Online]. Available: https://ec.europa.eu/eurostat/statistics-explained/index.php/Cancer_statistics#Deaths_from_cancer [Accessed 04/01/2019].
- FANG, C., OTERO, H. J., GREENBERG, D. & NEUMANN, P. J. 2011. Cost-utility analyses of diagnostic laboratory tests: a systematic review. *Value in Health*, 14, 1010-1018.
- FARAHANI, N. & PANTANOWITZ, L. 2015. Overview of Telepathology. *Surgical Pathology Clinics*.
- FARAHANI, N., PARWANI, A. & PANTANOWITZ, L. 2015. Whole slide imaging in pathology: advantages, limitations, and emerging perspectives. *Pathol. Lab Med. Int*, 7, 23-33.
- FOOT, C. & HARRISON, T. 2011. *How to Improve Cancer Outcomes: Explaining England's Relatively Poor Cancer Survival Rates*, King's Fund.
- FORBES, L., SIMON, A., WARBURTON, F., BONIFACE, D., BRAIN, K. E., DESSAIX, A., DONNELLY, C., HAYNES, K., HVIDBERG, L. & LAGERLUND, M. 2013. Differences in cancer awareness and beliefs between Australia, Canada, Denmark, Norway, Sweden and the UK (the International Cancer Benchmarking Partnership): do they contribute to differences in cancer survival? *British Journal of Cancer*, 108, 292.

- FURRER, D., SANSCHAGRIN, F., JACOB, S. & DIORIO, C. 2015. Advantages and disadvantages of technologies for HER2 testing in breast cancer specimens. *American journal of clinical pathology*, 144, 686-703.
- FURUKAWA, T. A. & GUYATT, G. H. 2006. Sources of bias in diagnostic accuracy studies and the diagnostic process. *Canadian Medical Association Journal*, 174, 481-482.
- GARBER, A. M. & PHELPS, C. E. 1997. Economic foundations of cost-effectiveness analysis. *Journal of health economics*, 16, 1-31.
- GARRISON, L. J., LALLA, D., BRAMMER, M., BABIGUMIRA, J., WANG, B. & PEREZ, E. 2013a. Assessing the potential cost-effectiveness of retesting IHC0, IHC1+, or FISH-negative early stage breast cancer patients for HER2 status (Provisional abstract). *Cancer* [Online], 119. Available: <http://onlinelibrary.wiley.com/o/cochrane/cleed/articles/NHSEED-22013025080/frame.html>.
- GARRISON, L. J., LALLA, D., BRAMMER, M., BABIGUMIRA, J. B., WANG, B. & PEREZ, E. A. 2013b. Assessing the potential cost-effectiveness of retesting IHC0, IHC1+, or FISH-negative early stage breast cancer patients for HER2 status. *Cancer*, 119, 3113-3122.
- GARRISON, L. P., BABIGUMIRA, J. B., MASAQUEL, A., WANG, B. C., LALLA, D. & BRAMMER, M. 2015. The lifetime economic burden of inaccurate HER2 testing: estimating the costs of false-positive and false-negative HER2 test results in US patients with early-stage breast cancer. *Value in Health*, 18, 541-546.
- GIULIANO, A. E., DALE, P. S., TURNER, R. R., MORTON, D. L., EVANS, S. W. & KRASNE, D. L. J. A. O. S. 1995. Improved axillary staging of breast cancer with sentinel lymphadenectomy. 222, 394.
- GOACHER, E., RANDELL, R., WILLIAMS, B., TREANOR, D. J. A. O. P. & MEDICINE, L. 2016. The diagnostic concordance of whole slide imaging and light microscopy: a systematic review. 141, 151-161.
- GOLDHIRSCH, A., GELBER, R. D., PICCART-GEHBART, M. J., DE AZAMBUJA, E., PROCTER, M., SUTER, T. M., JACKISCH, C., CAMERON, D., WEBER, H. A. & HEINZMANN, D. 2013. 2 years versus 1 year of adjuvant trastuzumab for HER2-positive breast cancer (HERA): an open-label, randomised controlled trial. *The Lancet*, 382, 1021-1028.
- GOV.UK. 2018. *Artificial Intelligence to help save lives at five new technology centres* [Online]. Available: <https://www.gov.uk/government/news/artificial-intelligence-to-help-save-lives-at-five-new-technology-centres> [Accessed 29/01/2019].
- GREEN, S. & PIEHL, J. J. J. O. P. 2003. Clinical practice guidelines: a guide to better practice, not a recipe for uniformity. 49, 3-4.
- GRIFFIN, J. & TREANOR, D. 2017. Digital pathology in clinical use: where are we now and what is holding us back? *Histopathology*, 70, 134-145.
- GURAY, M. & SAHIN, A. A. 2006. Benign breast diseases: classification, diagnosis, and management. *The oncologist*, 11, 435-449.
- HAIDICH, A.-B. 2010. Meta-analysis in medical research. *Hippokratia*, 14, 29.
- HAMMOND, M. E. H., HAYES, D. F., DOWSETT, M., ALLRED, D. C., HAGERTY, K. L., BADVE, S., FITZGIBBONS, P. L., FRANCIS, G., GOLDSTEIN, N. S., HAYES, M., HICKS, D. G., LESTER, S., LOVE, R., MANGU, P. B., MCSHANE, L., MILLER, K., OSBORNE, C. K., PAIK, S., PERLMUTTER, J., RHODES, A., SASANO, H., SCHWARTZ, J. N., SWEEP, F. C. G., TAUBE, S., TORLAKOVIC, E. E., VALENSTEIN, P., VIALE, G., VISSCHER, D., WHEELER, T., WILLIAMS, R. B., WITTLIFF, J. L. & WOLFF, A. C. 2010. American society of clinical oncology/college of american pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *Journal of Clinical Oncology*, 28, 2784-2795.

- HANSEN, R. P., VEDSTED, P., SOKOLOWSKI, I., SØNDERGAARD, J. & OLESEN, F. 2011. Time intervals from first symptom to treatment of cancer: a cohort study of 2,212 newly diagnosed cancer patients. *BMC health services research*, 11, 284.
- HAVRILESKY, L. J., CHINO, J. P. & MYERS, E. R. 2013. How much is another randomized trial of lymph node dissection in endometrial cancer worth? A value of information analysis. *Gynecologic Oncology*, 131, 140-146.
- HEALTH, N. I. F. & EXCELLENCE, C. 2009. *Early and locally advanced breast cancer: diagnosis and treatment*, National Institute for Health and Clinical Excellence.
- HEERDT, A. S. J. J. O. 2018. Lymphatic Mapping and Sentinel Lymph Node Biopsy for Breast Cancer. 4, 431-431.
- HELIN, H. O., TUOMINEN, V. J., YLINEN, O., HELIN, H. J. & ISOLA, J. 2016. Free digital image analysis software helps to resolve equivocal scores in HER2 immunohistochemistry. *Virchows Archiv*, 468, 191-198.
- HICKS, D. G. & SCHIFFHAUER, L. 2011. Standardized assessment of the HER2 status in breast cancer by immunohistochemistry. *Laboratory Medicine*, 42, 459-467.
- HIOM, S. 2015. Diagnosing cancer earlier: reviewing the evidence for improving cancer survival. *British Journal of Cancer*, 112, S1-S5.
- HO, J., AHLERS, S. M., STRATMAN, C., ARIDOR, O., PANTANOWITZ, L., FINE, J. L., KUZMISHIN, J. A., MONTALTO, M. C. & PARWANI, A. V. 2014. Can digital pathology result in cost savings? A financial projection for digital pathology implementation at a large integrated health care organization. *Journal of pathology informatics*, 5.
- HOAD, K. 2016. 'Introduction' [PowerPoint Presentation]. *IB2110 Simulation*: Warwick Business School, 13/01/2016.
- HOLTEN-ROSSING, H., MOLLER TALMAN, M. L., KRISTENSSON, M. & VAINER, B. 2015. Optimizing HER2 assessment in breast cancer: application of automated image analysis. *Breast Cancer Research and Treatment*, 152, 367-375.
- HOLTEN-ROSSING, H., TALMAN, M. L. M., JYLLING, A. M. B., LÆNKHOLM, A. V., KRISTENSSON, M. & VAINER, B. J. H. 2017. Application of automated image analysis reduces the workload of manual screening of sentinel lymph node biopsies in breast cancer. 71, 866-873.
- HOOMANS, T. & SEVERENS, J. L. 2014. Economic evaluation of implementation strategies in health care. BioMed Central.
- HORNBERGER, J., ALVARADO, M. D., REBECCA, C., GUTIERREZ, H. R., YU, T. M. & GRADISHAR, W. J. 2012. Clinical validity/utility, change in practice patterns, and economic implications of risk stratifiers to predict outcomes for early-stage breast cancer: a systematic review. *Journal of the National Cancer Institute*, 104, 1068-1079.
- HUSEREAU, D., DRUMMOND, M., PETROU, S., CARSWELL, C., MOHER, D., GREENBERG, D., AUGUSTOVSKI, F., BRIGGS, A. H., MAUSKOPF, J., LODER, E. J. C. E. & ALLOCATION, R. 2013. Consolidated health economic evaluation reporting standards (CHEERS) statement. 11, 6.
- HUXLEY, N., JONES-HUGHES, T., COELHO, H., SNOWSILL, T., COOPER, C., MENG, Y., HYDE, C. & MUJICA MOTA, R. 2015. *A systematic review and economic evaluation of intraoperative tests [RD-100i one-step nucleic acid amplification (OSNA) system and Metasin test] for detecting sentinel lymph node metastases in breast cancer*.
- ICR. 2017. HERA [Online]. Available: <https://www.icr.ac.uk/our-research/our-research-centres/clinical-trials-and-statistics-unit/clinical-trials/hera> [Accessed 06/10/2017].
- INDEPENDENT CANCER TASKFORCE. 2015. *Achieving world-class cancer outcomes. A strategy for england 2015-2020* [Online]. NHS England [Accessed 12/10/2018].
- IRWIG, L., MACASKILL, P., GLASZIOU, P. & FAHEY, M. 1995. Meta-analytic methods for diagnostic test accuracy. *Journal of clinical epidemiology*, 48, 119-130.

- JACKSON, S. E. & CHESTER, J. D. J. I. J. O. C. 2015. Personalised cancer medicine. 137, 262-266.
- JACOBSON, S. H., HALL, S. N. & SWISHER, J. R. 2006. Discrete-event simulation of health care systems. *Patient flow: Reducing delay in healthcare delivery*. Springer.
- JADPRO. 2016. *Managing the Care of Patients With HER2-Positive Breast Cancer: A Collaborative Practice Model* [Online]. Available: <http://www.advancedpractitioner.com/narratives/her2/> [Accessed 27/09/2017].
- JOENSUU, H., BONO, P., KATAJA, V., ALANKO, T., KOKKO, R., ASOLA, R., UTRIAINEN, T., TURPEENNIEMI-HUJANEN, T., JYRKIÖ, S. & MÖYKKYNNEN, K. 2009. Fluorouracil, epirubicin, and cyclophosphamide with either docetaxel or vinorelbine, with or without trastuzumab, as adjuvant treatments of breast cancer: final results of the FinHer Trial. *Journal of Clinical Oncology*, 27, 5685-5692.
- JOINT FORMULARY COMMITTEE 2017. *British national formulary (BNF) 73*, Pharmaceutical Press.
- JOSHI, A. S., SHARANGPANI, G. M., PORTER, K., KEYHANI, S., MORRISON, C., BASU, A. S., GHOLAP, G. A., GHOLAP, A. S. & BARSKY, S. H. 2007. Semi-automated imaging system to quantitate Her-2/neu membrane receptor immunoreactivity in human breast cancer. *Cytometry Part A*, 71, 273-285.
- JUDD, R., GIBSON, J. & GRAY, D. J. 2001. Cost effective algorithm for optimum utilization of IHC, FISH and image analysis for HER-2 testing in breast cancer. *Laboratory Investigation*, 81, 28A-28A.
- KANE, B. & LUZ, S. 2013. "Do no harm": Fortifying MDT collaboration in changing technological times. *International journal of medical informatics*, 82, 613-625.
- KANE, B., LUZ, S., O'BRIAN, D. S. & MCDERMOTT, R. 2007. Multidisciplinary team meetings and their impact on workflow in radiology and pathology departments. *BMC medicine*, 5, 15.
- KAPILA, S. N., BOAZ, K., NATARAJAN, S. J. I. J. O. A. & RESEARCH, B. M. 2016. The post-analytical phase of histopathology practice: Storage, retention and use of human tissue specimens. 6, 3.
- KARAKUSEVIC, S., EDWARDS, N., LEWIS, R. & DAYAN, M. 2016. The future of pathology services. *Nuffield Trust*.
- KARNON, J., STAHL, J., BRENNAN, A., CARO, J. J., MAR, J. & MÖLLER, J. J. M. D. M. 2012. Modeling using discrete event simulation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-4. 32, 701-711.
- KATZORKE, N., RACK, B., HAEBERLE, L., NEUGEBAUER, J., MELCHER, C., HAGENBECK, C., FORSTBAUER, H., ULMER, H. U., SOELING, U. & KREIENBERG, R. 2013. Prognostic value of HER2 on breast cancer survival. *J Clin Oncol*.
- KEAY, T., CONWAY, C., O'FLAHERTY, N., HEWITT, S. M., SHEA, K. & GAVRIELIDES, M. A. 2013. Reproducibility in the automated quantitative assessment of HER2/neu for breast cancer. *Journal of pathology informatics*, 4, 19.
- KEEFE, D. L. 2002. Trastuzumab-associated cardiotoxicity. *Cancer*, 95, 1592-1600.
- KENNEDY, P. J., LEATHLEY, C. M. & HUGHES, C. F. J. M. J. O. A. 2010. Clinical practice variation. 193, S97.
- KHAZAI, L., MIDDLETON, L. P., GOKTEPE, N., LIU, B. T. & SAHIN, A. A. 2015. Breast pathology second review identifies clinically significant discrepancies in over 10% of patients. *Journal of surgical oncology*, 111, 192-197.
- KLOK, R. M. & POSTMA, M. J. 2004. Four quadrants of the cost-effectiveness plane: some considerations on the south-west quadrant. *Expert Review of Pharmacoeconomics & Outcomes Research*, 4, 599-601.

- KUFE, D. W., POLLOCK, R. E., WEICHSELBAUM, R. R., BAST, R. C., GANSLER, T. S., HOLLAND, J. F., FREI, E., CONNOLLY, J. L., SCHNITT, S. J. & WANG, H. H. 2003. Role of the surgical pathologist in the diagnosis and management of the cancer patient.
- KUMAR, R. K., VELAN, G. M., KORELL, S. O., KANDARA, M., DEE, F. R. & WAKEFIELD, D. 2004. Virtual microscopy for learning and assessment in pathology. *The Journal of pathology*, 204, 613-618.
- KUNKLER, I., CANNEY, P., VAN TIENHOVEN, G. & RUSSELL, N. J. C. O. 2008. Elucidating the role of chest wall irradiation in 'intermediate-risk' breast cancer: the MRC/EORTC SUPREMO trial. 20, 31-34.
- KUNTZ, K., SAINFORT, F., BUTLER, M., TAYLOR, B., KULASINGAM, S., GREGORY, S., MANN, E., ANDERSON, J. M. & KANE, R. L. 2013. Decision and Simulation Modeling Alongside Systematic Reviews.
- LALKHEN, A. G. & MCCLUSKEY, A. 2008. Clinical tests: sensitivity and specificity. *Continuing Education in Anaesthesia Critical Care Pain*, 8, 221-223.
- LAUDICELLA, M., WALSH, B., BURNS, E., DONNI, P. L. & SMITH, P. C. 2018. What is the impact of rerouting a cancer diagnosis from emergency presentation to GP referral on resource use and survival? Evidence from a population-based study. *BMC cancer*, 18, 394.
- LAURINAVICIENE, A., DASEVICIUS, D., OSTAPENKO, V., JARMALAITES, S., LAZUTKA, J. & LAURINAVICIUS, A. 2011. Membrane connectivity estimated by digital image analysis of HER2 immunohistochemistry is concordant with visual scoring and fluorescence in situ hybridization results: algorithm evaluation on breast cancer tissue microarrays. *Diagnostic pathology*, 6, 87.
- LAW, A. M. Designing a simulation study: how to conduct a successful simulation study. Proceedings of the 35th conference on Winter simulation: driving innovation, 2003. Winter Simulation Conference, 66-70.
- LEE, A., CARDER, P., DEB, R., HOWE, M., KNOX, F., SHRIMANKAR, J., ANDERSON, N., COOKE, J., ELLIS, I. & JENKINS, J. J. L., UK: THE ROYAL COLLEGE OF PATHOLOGISTS 2016. Guidelines for non-operative diagnostic procedures and reporting in breast cancer screening.
- LEE, J. A., SHAHEEN, M., WALKE, T. & DALY, M. 2011. Clinical and health economic outcomes of alternative HER2 test strategies for guiding adjuvant trastuzumab therapy. *Expert Review of Pharmacoeconomics & Outcomes Research*, 11, 325-341.
- LEONG, A. S. Y. & ZHUANG, Z. P. 2011. The Changing Role of Pathology in Breast Cancer Diagnosis and Treatment. *PATHOBIOLOGY*, 78, 99-114.
- LILFORD, R. J., CHILTON, P. J., HEMMING, K., GIRLING, A. J., TAYLOR, C. A. & BARACH, P. 2010. Evaluating policy and service interventions: framework to guide selection and interpretation of study end points. *Bmj*, 341, c4413.
- LLOYD, M. C., ALLAM-NANDYALA, P., PUROHIT, C. N., BURKE, N., COPPOLA, D. & BUI, M. M. 2010. Using image analysis as a tool for assessment of prognostic and predictive biomarkers for breast cancer: how reliable is it? *Journal of pathology informatics*, 1, 29.
- LOHR, K. N. & FIELD, M. J. 1990. *Clinical practice guidelines: directions for a new program*, National Academies Press.
- LONG, T. H., LAWCE, H., DURUM, C., MOORE, S. R., OLSON, S. B., GATTER, K. & TROXELL, M. L. 2015. The new equivocal: changes to HER2 FISH results when applying the 2013 ASCO/CAP guidelines. *American journal of clinical pathology*, 144, 253-262.
- LONGACRE, T. A., ENNIS, M., QUENNEVILLE, L. A., BANE, A. L., BLEIWEISS, I. J., CARTER, B. A., CATELANO, E., HENDRICKSON, M. R., HIBSHOOSH, H. & LAYFIELD, L. J. 2006. Interobserver agreement and reproducibility in classification of invasive breast carcinoma: an NCI breast cancer family registry study. *Modern pathology*, 19, 195.

- LOOK HONG, N. J., CLARKE, G. M., YAFFE, M. J. & HOLLOWAY, C. M. B. 2016a. Cost-effectiveness analysis of whole-mount pathology processing for patients with early breast cancer undergoing breast conservation. *Current Oncology*, 23, S23-S31.
- LOOK HONG, N. J., CLARKE, G. M., YAFFE, M. J. & HOLLOWAY, C. M. B. 2016b. The cost of accuracy: A budget impact analysis of whole-mount histopathology processing for patients with breast cancer undergoing breast conservation. *Journal of Cancer Policy*, 9, 1-7.
- LORD CARTER, O. C. 2006. Report of the review of NHS pathology services in England—chaired by Lord Carter of Coles *London: National Health Services*.
- LORD CARTER, O. C. 2008. Report of the second phase of the review of NHS pathology services in England. *London: National Health Services*.
- LU, B. & GATSONIS, C. J. S. I. M. 2013. Efficiency of study designs in diagnostic randomized clinical trials. 32, 1451-1466.
- LUGTENBERG, M., BURGERS, J., WESTERT, G. J. B. Q. & SAFETY 2009. Effects of evidence-based clinical practice guidelines on quality of care: a systematic review. 18, 385-392.
- LYMAN, G. H., GIULIANO, A. E., SOMERFIELD, M. R., BENSON III, A. B., BODURKA, D. C., BURSTEIN, H. J., COCHRAN, A. J., CODY III, H. S., EDGE, S. B. & GALPER, S. J. J. O. C. O. 2005. American Society of Clinical Oncology guideline recommendations for sentinel lymph node biopsy in early-stage breast cancer. 23, 7703-7720.
- MA, X., SURI, M. F. K. & CHU, H. 2014. A trivariate meta-analysis of diagnostic studies accounting for prevalence and non-evaluable subjects: re-evaluation of the meta-analysis of coronary CT angiography studies. *BMC medical research methodology*, 14, 128.
- MAASKANT-BRAAT, A. J., VAN DE POLL-FRANSE, L. V., VOOGD, A. C., COEBERGH, J. W. W., ROUMEN, R. M., NOLTHENIUS-PUYLAERT, M. C. T., NIEUWENHUIZEN, G. A. J. B. C. R. & TREATMENT 2011. Sentinel node micrometastases in breast cancer do not affect prognosis: a population-based study. 127, 195-203.
- MACASKILL, P., GATSONIS, C., DEEKS, J., HARBORD, R. & TAKWOINGI, Y. 2010. Cochrane handbook for systematic reviews of diagnostic test accuracy. *Version 0.9. 0. London: The Cochrane Collaboration*.
- MACMILLAN. 2016. *Understanding breast cancer* [Online]. Available: <https://www.macmillan.org.uk/information-and-support/breast-cancer/understanding-cancer> [Accessed 27/09/2017].
- MADABHUSHI, A. 2009. Digital pathology image analysis: opportunities and challenges.
- MADABHUSHI, A. & LEE, G. 2016. Image analysis and machine learning in digital pathology: Challenges and opportunities. Elsevier.
- MAHFOUZ, S. M. 2009. OVERVIEW OF PATHOLOGY AND ITS RELATED DISCIPLINES. *Medical Sciences*, 1, 336.
- MAJEED, A., ALLWOOD, D., FOLEY, K. & BINDMAN, A. 2018. Healthcare outcomes and quality in the NHS: how do we compare and how might the NHS improve? *BMJ*, 362, k3036.
- MANSOURNIA, M. A., HIGGINS, J. P., STERNE, J. A. & HERNÁN, M. A. J. E. 2017. Biases in randomized trials: a conversation between trialists and epidemiologists. 28, 54.
- MARAS, G. 2015. Digital Pathology in Primary Diagnosis. *Hospital Healthcare Europe*, 191-3.
- MARINCE, C., WALTERS, S., RACHET, B., BUTLER, J., FIELDS, T., FINAN, P., MAXWELL, R., NEDREBØ, B., PÅHLMAN, L. & SJÖVALL, A. 2013. Stage at diagnosis and colorectal cancer survival in six high-income countries: a population-based study of patients diagnosed during 2000–2007. *Acta oncologica*, 52, 919-932.
- MASOOD, M. & HORSFIELD, C. J. B. 2016. How to succeed in your histopathology application. 355, i6301.

- MAUSKOPF, J. A., PAUL, J. E., GRANT, D. M. & STERGACHIS, A. J. P. 1998. The role of cost—consequence analysis in healthcare decision—making. 13, 277-288.
- MAYOR, S. 2006. NICE approves trastuzumab for early stage breast cancer. *BMJ: British Medical Journal*, 332, 1409.
- MCCABE, C., CLAXTON, K. & CULYER, A. J. 2008. The NICE cost-effectiveness threshold. *Pharmacoeconomics*, 26, 733-744.
- MCPHAIL, S., ELLISS-BROOKES, L., SHELTON, J., IVES, A., GREENSLADE, M., VERNON, S., MORRIS, E. & RICHARDS, M. 2013. Emergency presentation of cancer and short-term mortality. *British journal of cancer*, 109, 2027-2034.
- MENKE, J. 2014. Bayesian bivariate meta-analysis of sensitivity and specificity: summary of quantitative findings in 50 meta-analyses. *Journal of evaluation in clinical practice*, 20, 844-852.
- MIEOG, J., VAN DER HAGE, J. & VAN DE VELDE, C. 2007. Neoadjuvant chemotherapy for operable breast cancer. *British Journal of Surgery*, 94, 1189-1200.
- MILLS, A. M., GRADECKI, S. E., HORTON, B. J., BLACKWELL, R., MOSKALUK, C. A., MANDELL, J. W., MILLS, S. E. & CATHRO, H. P. 2017. Diagnostic Efficiency in Digital Pathology: A Comparison of Optical Versus Digital Assessment in 510 Surgical Pathology Cases. *The American Journal of Surgical Pathology*.
- MILTENBURG, D. M., MILLER, C., KARAMLOU, T. B. & BRUNICARDI, F. C. J. J. O. S. R. 1999. Meta-analysis of sentinel lymph node biopsy in breast cancer. 84, 138-142.
- MINOT, D., VOSS, J., RADEMACHER, S., LWIN, T., ORSULAK, J., NASSAR, A., CHEN, B. & CLAYTON, A. 2010. Automated Cellular Imaging System (ACIS III) for HER2 analysis—reproducibility and concordance with fluorescence in situ hybridization (FISH) of a laboratory validated scoring technique. *Cancer Cytopathology*, 5), 316-317.
- MINOT, D. M., KIPP, B. R., ROOT, R. M., MEYER, R. G., REYNOLDS, C. A., NASSAR, A., HENRY, M. R. & CLAYTON, A. C. 2009. Automated Cellular Imaging System III for Assessing HER2 Status in Breast Cancer Specimens. *American Journal of Clinical Pathology*, 132, 133-138.
- MINOT, D. M., VOSS, J., RADEMACHER, S., LWIN, T., ORSULAK, J., CARON, B., KETTERLING, R., NASSAR, A., CHEN, B. & CLAYTON, A. 2012. Image analysis of HER2 immunohistochemical staining: Reproducibility and concordance with fluorescence in situ hybridization of a laboratory-validated scoring technique. *American Journal of Clinical Pathology*, 137, 270-276.
- MISIALEK, M. J. 2014. Valuing value: the changing role of pathologists. Oxford University Press Oxford, UK.
- MITRI, Z., CONSTANTINE, T. & O'REGAN, R. 2012. The HER2 receptor in breast cancer: pathophysiology, clinical use, and new advances in therapy. *Chemotherapy research and practice*, 2012.
- MOHAMMED, Z., GOING, J. J., MCMILLAN, D. C., ORANGE, C., MALLON, E., DOUGHTY, J. C. & EDWARDS, J. 2012. Comparison of visual and automated assessment of HER2 status and their impact on outcome in primary operable invasive ductal breast cancer. *Histopathology*, 61, 675-684.
- MOHER, D., SHAMSEER, L., CLARKE, M., GHERSI, D., LIBERATI, A., PETTICREW, M., SHEKELLE, P. & STEWART, L. A. 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic reviews*, 4, 1.
- MONTGOMERY, M. & MCCRONE, S. H. 2010. Psychological distress associated with the diagnostic phase for suspected breast cancer: systematic review. *Journal of advanced nursing*, 66, 2372-2390.
- MORELLE, M., HASLÉ, E., TREILLEUX, I., MICHOT, J.-P., BACHELOT, T., PENAULT-LLORCA, F. & CARRÈRE, M.-O. 2006. Cost-effectiveness analysis of strategies for HER2 testing

- of breast cancer patients in France. *International journal of technology assessment in health care*, 22, 396-401.
- MORGAN, I. & WILKES, S. 2017. Improving early diagnosis of cancer in UK general practice. *British Journal of General Practice*, 67, 276-277.
- MUGHAL, M. & GOODMAN, J. 2017. MDT IMPROVEMENT REPORT.
- MULLER, P., WALTERS, S., COLEMAN, M. P. & WOODS, L. 2018. Which indicators of early cancer diagnosis from population-based data sources are associated with short-term mortality and survival? *Cancer epidemiology*, 56, 161-170.
- MURAD, M. H., ASI, N., ALSAWAS, M. & ALAHDAB, F. 2016. New evidence pyramid. *BMJ Evidence-Based Medicine*, ebmed-2016-110401.
- MURCHIE, P., SMITH, S. M., YULE, M. S., ADAM, R., TURNER, M. E., LEE, A. J. & FIELDING, S. 2017. Does emergency presentation of cancer represent poor performance in primary care? Insights from a novel analysis of linked primary and secondary care data. *British journal of cancer*, 116, 1148-1158.
- NASSAR, A., COHEN, C., AGERSBORG, S. S., ZHOU, W., LYNCH, K. A., ALBITAR, M., BARKER, E. A., VANDERBILT, B. L., THOMPSON, J. & HEYMAN, E. R. 2011a. Trainable immunohistochemical HER2/neu image analysis: a multisite performance study using 260 breast tissue specimens. *Archives of pathology & laboratory medicine*, 135, 896-902.
- NASSAR, A., COHEN, C., AGERSBORG, S. S., ZHOU, W., LYNCH, K. A., HEYMAN, E. R., OLSON, A., LANGE, H., SIDDIQUI, M. T. J. A. I. & MORPHOLOGY, M. 2011b. A new immunohistochemical ER/PR image analysis system: a multisite performance study. 19, 195-202.
- NATIONAL CANCER ACTION TEAM. 2010. *The Characteristics of an Effective Multidisciplinary Team (MDT)* [Online]. Available: file:///C:/Users/u1111322/Downloads/NCATMDTCharacteristics_FINAL%20(1).pdf [Accessed 20/11/2017].
- NATIONAL CANCER INTELLIGENCE NETWORK. 2006. *All Breast Cancer Report. A UK analysis of all symptomatic and screen-detected breast cancers diagnosed in 2006* [Online]. Available: file:///C:/Users/u1111322/Downloads/091019_wmciu_all_breast_cancer_report%20(6).pdf [Accessed].
- NCIN. 2013. *Routes to diagnosis: comparing multiple studies* [Online]. Available: http://www.ncin.org.uk/publications/data_briefings/routes_to_diagnosis_comparing_multiple_studies [Accessed 05/01/2019].
- NEAL, R. D., DIN, N. U., HAMILTON, W., UKOUMUNNE, O. C., CARTER, B., STAPLEY, S. & RUBIN, G. 2014. Comparison of cancer diagnostic intervals before and after implementation of NICE guidelines: analysis of data from the UK General Practice Research Database. *British Journal of Cancer*, 110, 584-592.
- NEAL, R. D., THARMANATHAN, P., FRANCE, B., DIN, N. U., COTTON, S., FALLON-FERGUSON, J., HAMILTON, W., HENDRY, A., HENDRY, M., LEWIS, R., MACLEOD, U., MITCHELL, E. D., PICKETT, M., RAI, T., SHAW, K., STUART, N., TØRRING, M. L., WILKINSON, C., WILLIAMS, B., WILLIAMS, N. & EMERY, J. 2015. Is increased time to diagnosis and treatment in symptomatic cancer associated with poorer outcomes? Systematic review. *British Journal of Cancer*.
- NHS BSP. 2007. *Reporting, recording and auditing B5 core biopsies with normal/benign surgery* [Online]. NHS Cancer Screening Programmes. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/442171/nhsbsp-gpg9.pdf [Accessed 27/09/2017].
- NHS BSP. 2009. *Quality Assurance Guidelines for Surgeons in Breast Cancer Screening* [Online]. NHS Cancer Screening Programmes Available:

- https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/465694/nhsbsp20.pdf [Accessed 27/09/2017].
- NHS BSP. 2011a. *Quality Assurance Guidelines for Breast Cancer Screening Radiology* [Online]. NHS Cancer Screening Programmes. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/470579/nhsbsp59_QA_radiology_uploaded_231015.pdf [Accessed 27/09/2017].
- NHS BSP. 2011b. *Quality Assurance Guidelines for Breast Pathology Services* [Online]. NHS Cancer Screening Programme. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/465683/nhsbsp02.pdf [Accessed 27/09/2017].
- NHS ENGLAND. 2014a. *five year forward view* [Online]. Available: <https://www.england.nhs.uk/wp-content/uploads/2014/10/5yfv-web.pdf> [Accessed 12/10/2018].
- NHS ENGLAND. 2014b. *National Pathology Programme Digital First: Clinical Transformation through Pathology Innovation* [Online]. Available: <https://www.england.nhs.uk/wp-content/uploads/2014/02/pathol-dig-first.pdf> [Accessed 18/10/2018].
- NICE. 2002. *Improving outcomes in breast cancer (CSG1)* [Online]. Available: <https://www.nice.org.uk/guidance/csg1/resources/improving-outcomes-in-breast-cancer-update-pdf-773371117> [Accessed 27/09/2017].
- NICE. 2006a. *Hormonal therapies for the adjuvant treatment of early oestrogen-receptor-positive breast cancer Technology appraisal guidance [TA112]* [Online]. Available: <https://www.nice.org.uk/guidance/ta112> [Accessed 27/09/2017].
- NICE. 2006b. *Trastuzumab for the adjuvant treatment of early-stage HER2-positive breast cancer. Technology appraisal guidance [TA107]* [Online]. Available: <https://www.nice.org.uk/guidance/ta107> [Accessed 27/09/2017].
- NICE. 2009. *Early and locally advanced breast cancer: diagnosis and treatment* [Online]. Available: https://www.nice.org.uk/guidance/cg80/chapter/Key-priorities-for-implementation#ftn.footnote_3 [Accessed].
- NICE. 2011a. *Breast Cancer (QS12)* [Online]. Available: <https://www.nice.org.uk/guidance/qs12/resources/breast-cancer-pdf-2098481951941> [Accessed 27/09/2017].
- NICE. 2011b. *Early and locally advanced breast cancer overview* [Online]. Available: [file:///C:/Users/u1111322/Downloads/early-and-locally-advanced-breast-cancer-early-and-locally-advanced-breast-cancer-overview%20\(1\).pdf](file:///C:/Users/u1111322/Downloads/early-and-locally-advanced-breast-cancer-early-and-locally-advanced-breast-cancer-overview%20(1).pdf) [Accessed 27/09/2017].
- NICE. 2014. *Early and locally advanced breast cancer: adjuvant therapy* [Online]. Available: <http://pathways.nice.org.uk/pathways/early-and-locally-advancedbreast-cancer> [Accessed 27/09/2017].
- NITTA, H., KELLY, B. D., ALLRED, C., JEWELL, S., BANKS, P., DENNIS, E. & GROGAN, T. M. 2016. The assessment of HER2 status in breast cancer: the past, the present, and the future. *Pathology international*, 66, 313-324.
- NORELL, C., ROBINSON, D., BUTLER, J. & HARRISON, S. 2018. Exploring Variations in the Content of Cancer-Specific Treatment Guidelines: An International Cancer Benchmarking Partnership (ICBP) Study. American Society of Clinical Oncology.
- NUFFIELD TRUST. 2018. *Cancer survival rates* [Online]. Available: <https://www.nuffieldtrust.org.uk/resource/cancer-survival-rates> [Accessed 05/01/2019].

- O'BRIEN, B. J., GERTSEN, K., WILLAN, A. R. & FAULKNER, A. 2002. Is there a kink in consumers' threshold value for cost-effectiveness in health care? *Health economics*, 11, 175-180.
- O'DOWD, A. 2015. GP cancer referral delays may explain UK's low cancer survival rates. *BMJ: British Medical Journal*, 350.
- O'MAHONY, J. F., NEWALL, A. T. & VAN ROSMALEN, J. J. P. 2015. Dealing with time in health economic evaluation: methodological issues and recommendations for practice. 33, 1255-1268.
- ONITILLO, A. A., ENGEL, J. M. & STANKOWSKI, R. V. 2014. Cardiovascular toxicity associated with adjuvant trastuzumab therapy: prevalence, patient characteristics, and risk factors. *Therapeutic advances in drug safety*, 5, 154-166.
- ONS. 2015. *Survival from cancer improving and more people being diagnosed* [Online]. Available: <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/survivalfromcancerimprovingandmorepeoplebeingdiagnosed/2015-07-17> [Accessed 04/01/2019].
- OOSTENBRINK, J. B., AL, M. J., OPPE, M. & RUTTEN-VAN MÖLKEN, M. P. J. V. I. H. 2008. Expected value of perfect information: an empirical example of reducing decision uncertainty by conducting additional research. 11, 1070-1080.
- PALSER, T. R., CROMWELL, D. A., HARDWICK, R. H., RILEY, S. A., GREENAWAY, K. & VAN DER MEULEN, J. H. 2013. Impact of route to diagnosis on treatment intent and 1-year survival in patients diagnosed with oesophagogastric cancer in England: a prospective cohort study. *BMJ open*, 3, e002129.
- PANDHARIPANDE, P. V., HARISINGHANI, M. G., OZANNE, E. M., SPECHT, M. C., HUR, C., LEE, J. M. & GAZELLE, G. S. 2008. Staging MR lymphangiography of the axilla for early breast cancer: cost-effectiveness analysis. *American Journal of Roentgenology*, 191, 1308-1319.
- PANJWANI, P., EPARI, S., KARPATE, A., SHIRSAT, H., RAJSEKHARAN, P., BASAK, R., SHET, T., CHINYOY, R., CHACKO, R. & GURSALE, S. 2010. Assessment of HER-2/neu status in breast cancer using fluorescence in situ hybridization & immunohistochemistry: Experience of a tertiary cancer referral centre in India.
- PANTANOWITZ, L., VALENSTEIN, P. N., EVANS, A. J., KAPLAN, K. J., PFEIFER, J. D., WILBUR, D. C., COLLINS, L. C. & COLGAN, T. J. 2011. Review of the current state of whole slide imaging in pathology. *Journal of pathology informatics*, 2.
- PESEK, S., ASHIKAGA, T., KRAG, L. E. & KRAG, D. J. W. J. O. S. 2012. The false-negative rate of sentinel node biopsy in patients with breast cancer: a meta-analysis. 36, 2239-2251.
- PETROU, S. & GRAY, A. J. B. 2011. Economic evaluation using decision analytical modelling: design, conduct, analysis, and reporting. 342, d1766.
- PHE. 2017. *NHS Breast Screening Programme: Consolidated standards* [Online]. Public Health England. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/612739/Breast_screening_consolidated_standards.pdf [Accessed 27/09/2017].
- PHILIPS, Z., BOJKE, L., SCULPHER, M., CLAXTON, K. & GOLDER, S. J. P. 2006. Good practice guidelines for decision-analytic modelling in health technology assessment. 24, 355-371.
- PICCART-GEHBART, M. J., PROCTER, M., LEYLAND-JONES, B., GOLDBIRSH, A., UNTCH, M., SMITH, I., GIANNI, L., BASELGA, J., BELL, R. & JACKISCH, C. 2005. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *New England Journal of Medicine*, 353, 1659-1672.

- PINDER, S. E. & REIS-FILHO, J. S. 2007. Non-operative breast pathology. *Journal of clinical pathology*, 60, 1297-1299.
- PSSRU. 2016. *Unit Costs of Health and Social care* [Online]. Available: <http://www.pssru.ac.uk/pub/uc/uc2016/full.pdf> [Accessed 06/10/2017].
- PUBLIC HEALTH ENGLAND. 2016. *Stage at diagnosis 2012-2014 and one-year cancer survival in England* [Online]. Available: file:///C:/Users/u1111322/Downloads/NCRAS_Data_briefing_stage_and_survival_v6.pdf [Accessed 06/01/2019].
- PURUSHOTHAM, A., PAIN, S., MILES, D. & HARNETT, A. J. T. L. O. 2001. Variations in treatment and survival in breast cancer. 2, 719-725.
- QUARESMA, M., COLEMAN, M. P. & RACHET, B. J. T. L. 2015. 40-year trends in an index of survival for all cancers combined and survival adjusted for age and sex for each cancer in England and Wales, 1971–2011: a population-based study. 385, 1206-1218.
- RAAB, S. S., GRZYBICKI, D. M., JANOSKY, J. E., ZARBO, R. J., MEIER, F. A., JENSEN, C. & GEYER, S. J. J. C. 2005. Clinical impact and frequency of anatomic pathology errors in cancer diagnoses. 104, 2205-2213.
- RAAB, S. S. & GRZYBICKI, D. M. J. C. A. C. J. F. C. 2010. Quality in cancer diagnosis. 60, 139-165.
- RAFIA, R., BRENNAN, A., MADAN, J., COLLINS, K., REED, M. W., LAWRENCE, G., ROBINSON, T., GREENBERG, D. & WYLD, L. 2016. Modeling the cost-effectiveness of alternative upper age limits for breast cancer screening in England and Wales. *Value in Health*, 19, 404-412.
- RAHMAN, M. & MOHAMMED, S. 2015. Breast cancer metastasis and the lymphatic system. *Oncology letters*, 10, 1233-1239.
- RAKHA, E. A., ALESKANDARANI, M., TOSS, M. S., GREEN, A. R., BALL, G., ELLIS, I. O. & DALTON, L. W. J. J. O. C. P. 2018a. Breast cancer histologic grading using digital microscopy: concordance and outcome association. *jclinpath*-2017-204979.
- RAKHA, E. A., ALESKANDARANY, M. A., TOSS, M. S., MONGAN, N., ELSAYED, M. E., GREEN, A. R., ELLIS, I. O. & DALTON, L. W. J. H. 2018b. Impact of breast cancer grade discordance on prediction of outcome.
- RAKHA, E. A., PINDER, S. E., BARTLETT, J. M., IBRAHIM, M., STARCZYNSKI, J., CARDER, P. J., PROVENZANO, E., HANBY, A., HALES, S. & LEE, A. H. 2014. Updated UK Recommendations for HER2 assessment in breast cancer. *Journal of clinical pathology*, *jclinpath*-2014-202571.
- RCPATH. 2018a. *Diagnostic digital pathology* [Online]. Available: <https://www.rcpath.org/profession/digital-pathology.html> [Accessed 29/09/2018].
- RCPATH. 2018b. *TNM classification of tumours of the breast (UICC TNM 8)* [Online]. Available: https://www.rcpath.org/resourceLibrary/searchResults/?resource_keyword=breast [Accessed 12/10/2018].
- RCPATH. 2018c. *What is pathology?* [Online]. The Royal College of Pathologists. Available: <https://www.rcpath.org/discover-pathology/what-is-pathology.html> [Accessed 06/11/2018].
- RCPATH. 2019. *Research and Innovation* [Online]. Available: <https://www.rcpath.org/profession/research-and-innovation.html> [Accessed 13/01/2019].
- RCR. 2016. *Postoperative radiotherapy for breast cancer: UK consensus statements* [Online]. Royal College of Radiologists. Available: https://www.rcr.ac.uk/system/files/publication/field_publication_files/bfco2016_breast-consensus-guidelines.pdf [Accessed 27/09/2017].

- REYES, C., IKPATT, O. F., NADJI, M. & COTE, R. J. J. O. P. I. 2014. Intra-observer reproducibility of whole slide imaging for the primary diagnosis of breast needle biopsies. 5.
- RHODES, A., JASANI, B., BALATON, A., BARNES, D. & MILLER, K. J. J. O. C. P. 2000a. Frequency of oestrogen and progesterone receptor positivity by immunohistochemical analysis in 7016 breast carcinomas: correlation with patient age, assay sensitivity, threshold value, and mammographic screening. 53, 688-696.
- RHODES, A., JASANI, B., BARNES, D. E., BOBROW, L. & MILLER, K. J. J. O. C. P. 2000b. Reliability of immunohistochemical demonstration of oestrogen receptors in routine practice: interlaboratory variance in the sensitivity of detection and evaluation of scoring systems. 53, 125-130.
- RIBER-HANSEN, R., VAINER, B. & STEINICHE, T. 2012. Digital image analysis: a review of reproducibility, stability and basic requirements for optimal results. *Apmis*, 120, 276-289.
- RILEY, R. D., JACKSON, D., SALANTI, G., BURKE, D. L., PRICE, M., KIRKHAM, J. & WHITE, I. R. 2017. Multivariate and network meta-analysis of multiple outcomes and multiple treatments: rationale, concepts, and examples. *bmj*, 358, j3932.
- ROBINSON, R. 1993a. Cost-benefit analysis. *BMJ*, 307, 924-926.
- ROBINSON, R. 1993b. Cost-effectiveness analysis. *Bmj*, 307, 793-795.
- ROBINSON, R. 1993c. Cost-utility analysis. *BMJ*, 307, 859-862.
- ROBINSON, R. 1993d. Costs and cost-minimisation analysis. *Bmj*, 307, 726-728.
- ROBINSON, R. 1993e. Economic evaluation and health care. What does it mean? *Bmj*, 307, 670-673.
- ROBINSON, S. Simulation model verification and validation: increasing the users' confidence. Proceedings of the 29th conference on Winter simulation, 1997. IEEE Computer Society, 53-59.
- ROBINSON, S. 2004. *Simulation: the practice of model development and use*, Wiley Chichester.
- ROBINSON, S. 2008. Conceptual modelling for simulation Part I: definition and requirements. *Journal of the operational research society*, 59, 278-290.
- RODGER, M., RAMSAY, T. & FERGUSON, D. J. T. 2012. Diagnostic randomized controlled trials: the final frontier. 13, 137.
- ROEVER, L. 2015. Types of Bias in Studies of Diagnostic Test Accuracy. *Evidence Based Medicine and Practice Open Access Journal: 1000e113*. doi, 10, 2.
- ROJO, M. G., GARCÍA, G. B., MATEOS, C. P., GARCÍA, J. G. & VICENTE, M. C. 2006. Critical comparison of 31 commercially available digital slide systems in pathology. *International journal of surgical pathology*, 14, 285-305.
- ROOPE, R. & AFOM, M. R. The importance of early diagnosis in cancer patients.
- ROSE, P. W., RUBIN, G., PERERA-SALAZAR, R., ALMBERG, S. S., BARISIC, A., DAWES, M., GRUNFELD, E., HART, N., NEAL, R. D. & PIROTTA, M. 2015. Explaining variation in cancer survival between 11 jurisdictions in the International Cancer Benchmarking Partnership: a primary care vignette survey. *BMJ Open*, 5, e007212.
- ROSS, J. S. J. C. C. 2011. Point: Fluorescence in situ hybridization is the preferred approach over immunohistochemistry for determining HER2 status. 57, 980-982.
- ROSTAS, J. W. & DYESS, D. L. 2011. Current operative management of breast cancer: an age of smaller resections and bigger cures. *International journal of breast cancer*, 2012.
- RUBIN, I. & YARDEN, Y. 2001. The basic biology of HER2. *Annals of Oncology*, 12, S3-S8.
- RUIZ-CASADO, A., ORTEGA, M. J., SORIA, A. & CEBOLLA, H. 2014. Clinical audit of multidisciplinary care at a medium-sized hospital in Spain. *World journal of surgical oncology*, 12, 53.

- RUTTER, C. M. & GATSONIS, C. A. 2001. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistics in medicine*, 20, 2865-2884.
- SACCO, J. J., BOTTEN, J., MACBETH, F., BAGUST, A. & CLARK, P. 2010. The average body surface area of adult cancer patients in the UK: a multicentre retrospective study. *PloS one*, 5, e8933.
- SACERDOTE, C., BORDON, R., PITARELLA, S., MANO, M. P., BALDI, I., CASELLA, D., DI CUONZO, D., FRIGERIO, A., MILANESIO, L. & MERLETTI, F. J. B. H. S. R. 2013. Compliance with clinical practice guidelines for breast cancer treatment: a population-based study of quality-of-care indicators in Italy. 13, 28.
- SARGENT, R. G. Verification and validation of simulation models. Simulation Conference (WSC), Proceedings of the 2009 Winter, 2009. IEEE, 162-176.
- SATTA, G. & EDMONSTONE, J. 2018. Consolidation of pathology services in England: have savings been achieved? *BMC health services research*, 18, 862.
- SCHMIDT, C. 2011. How do you tell whether a breast cancer is HER2 positive? Ongoing studies keep debate in high gear. Oxford University Press.
- SCHUETZ, G. M., SCHLATTMANN, P. & DEWEY, M. 2012. Use of 3× 2 tables with an intention to diagnose approach to assess clinical performance of diagnostic tests: meta-analytical evaluation of coronary CT angiography studies. *Bmj*, 345, e6717.
- SCHWARZER, G. 2007. meta: An R package for meta-analysis. *R news*, 7, 40-45.
- SECTRA. 2015. *Digital pathology reduces turnaround time and barriers in using second opinions—thus improving patient care* [Online]. Available: <https://sectra.com/medical/resources/digital-pathology-reduces-turnaround-time-and-barriers-in-using-second-opinions/> [Accessed 17/11/2017].
- SHAH, M. V., WIKTOR, A. E., MEYER, R. G., TENNER, K. S., BALLMAN, K. V., KETTERLING, R. P., PEREZ, E. A. & JENKINS, R. B. 2014. Changing pattern for HER2 positivity due to updated ASCO/CAP guidelines for HER2 testing and its impact. American Society of Clinical Oncology.
- SIMUL8. 2019. *Simul8 software* [Online]. Available: <https://store.simul8.com/> [Accessed].
- SKALAND, I., OVESTAD, I., JANSSEN, E. A., KLOS, J., KJELLEVOLD, K. H., HELLIESEN, T. & BAAK, J. P. 2008. Digital image analysis improves the quality of subjective HER-2 expression scoring in breast cancer. *Applied Immunohistochemistry & Molecular Morphology*, 16, 185-90.
- SLAMON, D., CLARK, G., WONG, S., LEVIN, W., ULLRICH, A. & MCGUIRE, W. 1987. Human breast cancer: correlation of relapse and. *Science*, 3798106, 235.
- SLAMON, D. J., LEYLAND-JONES, B., SHAK, S., FUCHS, H., PATON, V., BAJAMONDE, A., FLEMING, T., EIERMANN, W., WOLTER, J. & PEGRAM, M. 2001. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *New England Journal of Medicine*, 344, 783-792.
- SLAOUI, M. & FIETTE, L. 2011. Histopathology procedures: from tissue sampling to histopathological evaluation. *Drug Safety Evaluation*. Springer.
- SNEAD, D. R., TSANG, Y. W., MESKIRI, A., KIMANI, P. K., CROSSMAN, R., RAJPOOT, N. M., BLESSING, E., CHEN, K., GOPALAKRISHNAN, K. & MATTHEWS, P. 2015. Validation of digital pathology imaging for primary histopathological diagnosis. *Histopathology*.
- SNEAD, D. R., TSANG, Y. W., MESKIRI, A., KIMANI, P. K., CROSSMAN, R., RAJPOOT, N. M., BLESSING, E., CHEN, K., GOPALAKRISHNAN, K. & MATTHEWS, P. J. H. 2016. Validation of digital pathology imaging for primary histopathological diagnosis. 68, 1063-1072.
- SRIGLEY, J. R. 2009. The pathologist as diagnostic oncologist. Taylor & Francis.
- STARFIELD, B., SHI, L. & MACINKO, J. 2005. Contribution of primary care to health systems and health. *The Milbank Quarterly*, 83, 457-502.

- STATHONIKOS, N., VETA, M., HUISMAN, A. & VAN DIEST, P. J. 2013. Going fully digital: Perspective of a Dutch academic pathology lab. *Journal of pathology informatics*, 4.
- STOSS, O. C., SCHEEL, A., NAGELMEIER, I., SCHILDHAUS, H.-U., HENKEL, T., VIALE, G., JASANI, B., UNTCH, M. & RÜSCHOFF, J. 2015. Impact of updated HER2 testing guidelines in breast cancer—re-evaluation of HERA trial fluorescence in situ hybridization data. *Modern Pathology*, 28, 1528.
- STRACHAN, C., HORGAN, K., MILLICAN-SLATER, R., SHAABAN, A. & SHARMA, N. 2016. Outcome of a new patient pathway for managing B3 breast lesions by vacuum-assisted biopsy: time to change current UK practice? *Journal of clinical pathology*, 69, 248-254.
- STRATMAN, C., DROGOWSKI, L. & HO, J. 2010. Digital pathology in the clinical workflow: A time and motion study. *Pathology Visions*.
- SUTTON, M., GARFIELD-BIRKBECK, S., MARTIN, G., MEACOCK, R., MORRIS, S., SCULPHER, M., STREET, A., WATSON, S. & LILFORD, R. 2018. Economic analysis of service and delivery interventions in health care. *Health Services and Delivery Research*.
- TAKWOINGI, Y., GUO, B., RILEY, R. D. & DEEKS, J. J. 2017. Performance of methods for meta-analysis of diagnostic test accuracy with few studies or sparse data. *Statistical methods in medical research*, 26, 1896-1911.
- TAKWOINGI, Y., RILEY, R. D. & DEEKS, J. J. 2015. Meta-analysis of diagnostic accuracy studies in mental health. *Evidence-based mental health*, ebmental-2015-102228.
- TANG, S. S.-K., KAPTANIS, S., HADDOW, J. B., MONDANI, G., ELSBERGER, B., TASOULIS, M. K., OBONDO, C., JOHNS, N., ISMAIL, W. & SYED, A. J. E. J. O. C. 2017. Current margin practice and effect on re-excision rates following the publication of the SSO-ASTRO consensus and ABS consensus guidelines: a national prospective study of 2858 women undergoing breast-conserving therapy in the UK and Ireland. 84, 315-324.
- TAYLOR, C. & RAMIREZ, A. 2009. Multidisciplinary team members' views about MDT working: results from a survey commissioned by the National Cancer Action Team. *London: NHS National Cancer Action Team*.
- TÊTU, B. & EVANS, A. 2013. Canadian licensure for the use of digital pathology for routine diagnoses: one more step toward a new era of pathology practice without borders. *Archives of Pathology and Laboratory Medicine*, 138, 302-304.
- THOMAS, L. J. E.-B. N. 1999. Clinical practice guidelines. 2, 38-39.
- THORN, J., COAST, J. & ANDRONIS, L. J. M. D. M. 2016. Interpretation of the expected value of perfect information and research recommendations: a systematic review and empirical investigation. 36, 285-295.
- THORSTENSON, S., MOLIN, J. & LUNDSTRÖM, C. 2014. Implementation of large-scale routine diagnostics using whole slide imaging in Sweden: Digital pathology experiences 2006-2013. *Journal of pathology informatics*, 5.
- TORRES, M. I., WEBSTER, F. B., BILOUS, A. M., REED, R. G., KOSSOFF, M. B., GODDING, R. M. & REDMAN, S. 2003. Supporting the implementation of breast pathology recommendations: feasibility of a feedback mechanism to pathologists. *Pathology*, 35, 402-8.
- TURASHVILI, G., LEUNG, S., TURBIN, D., MONTGOMERY, K., GILKS, B., WEST, R., CARRIER, M., HUNTSMAN, D. & APARICIO, S. 2009. Inter-observer reproducibility of HER2 immunohistochemical assessment and concordance with fluorescent in situ hybridization (FISH): pathologist assessment compared to quantitative image analysis. *BMC cancer*, 9, 165.
- TURBIN, D. A., LEUNG, S., CHEANG, M. C., KENNECKE, H. A., MONTGOMERY, K. D., MCKINNEY, S., TREABA, D. O., BOYD, N., GOLDSTEIN, L. C., BADVE, S. J. B. C. R. & TREATMENT 2008. Automated quantitative analysis of estrogen receptor

- expression in breast carcinoma does not differ from expert pathologist scoring: a tissue microarray study of 3,484 cases. 110, 417-426.
- ULANER, G. A., RIEDL, C. C., DICKLER, M. N., JHAVERI, K., PANDIT-TASKAR, N. & WEBER, W. J. J. O. N. M. O. P., SOCIETY OF NUCLEAR MEDICINE 2016. Molecular imaging of biomarkers in breast cancer. 57, 53S.
- VARGA, Z. & NOSKE, A. 2015. Impact of modified 2013 ASCO/CAP guidelines on HER2 testing in breast cancer. One year experience. *PLoS One*, 10, e0140652.
- VODOVNIK, A. 2015. Distance reporting in digital pathology: A study on 950 cases. *Journal of pathology informatics*, 6.
- VODOVNIK, A. 2016. Diagnostic time in digital pathology: A comparative study on 400 cases. *Journal of pathology informatics*, 7.
- WALKER, E., HERNANDEZ, A. V. & KATTAN, M. W. J. C. C. J. O. M. 2008. Meta-analysis: Its strengths and limitations. 75, 431.
- WALKER, R. 2006. Quantification of immunohistochemistry—issues concerning methods, utility and semiquantitative assessment I. *Histopathology*, 49, 406-410.
- WALTERS, S., MARINGE, C., COLEMAN, M. P., PEAKE, M. D., BUTLER, J., YOUNG, N., BERGSTRÖM, S., HANNA, L., JAKOBSEN, E. & KÖLBECK, K. 2013. Lung cancer survival and stage at diagnosis in Australia, Canada, Denmark, Norway, Sweden and the UK: a population-based study, 2004–2007. *Thorax*, 68, 551-564.
- WANG, S., SABOORIAN, M., FRENKEL, E., HALEY, B. & SIDDIQUI, M. 2001. Automated Cellular Imaging System (ACIS)—assisted quantitation of immunohistochemical assay achieves high accuracy in comparison with fluorescence in situ hybridization assay as the standard. *Am J Clin Pathol*, 116, 495-503.
- WATTS, R. D., LI, I. W., GEELHOED, E. A., SANFILIPPO, F. M. & JOHN, A. S. 2017. Economic Evaluations of Pathology Tests, 2010-2015: A Scoping Review. *Value in Health*, 20, 1210-1215.
- WEBSTER, J. & DUNSTAN, R. 2014. Whole-slide imaging and automated image analysis: considerations and opportunities in the practice of pathology. *Veterinary pathology*, 51, 211-223.
- WEINSTEIN, R. S., DESOUR, M. R., LIANG, C., BHATTACHARYYA, A. K., GRAHAM, A. R., DAVIS, J. R., SCOTT, K. M., RICHTER, L., KRUPINSKI, E. A., SZYMUS, J., KAYSER, K. & DUNN, B. E. 2001. Telepathology overview: From concept to implementation. *Human Pathology*, 32, 1283-1299.
- WEINSTEIN, R. S., GRAHAM, A. R., RICHTER, L. C., BARKER, G. P., KRUPINSKI, E. A., LOPEZ, A. M., ERPS, K. A., BHATTACHARYYA, A. K., YAGI, Y. & GILBERTSON, J. R. 2009. Overview of telepathology, virtual microscopy, and whole slide imaging: prospects for the future. *Human pathology*, 40, 1057-1069.
- WELTON, N., SUTTON, A., COOPER, N., ABRAMS, K. & ADES, A. 2012. Evidence Synthesis in a Decision Modelling Framework. *Evidence synthesis for decision making in healthcare*. John Wiley & Sons Ltd.
- WHITING, P. F., RUTJES, A. W., WESTWOOD, M. E., MALLETT, S., DEEKS, J. J., REITSMA, J. B., LEEFLANG, M. M., STERNE, J. A. & BOSSUYT, P. M. 2011. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of internal medicine*, 155, 529-536.
- WHO. 2018. *Cancer: key facts* [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/cancer> [Accessed 05/01/2019].
- WILLETT, A., MICHELL, M. & LEE, M. J. A. A. T. N. 2010. Best practice diagnostic guidelines for patients presenting with breast symptoms. 2010.
- WILLIAMS, B. J., BOTTOMS, D. & TREANOR, D. 2017. Future-proofing pathology: the case for clinical adoption of digital pathology. *Journal of Clinical Pathology*, jclinpath-2017-204644.

- WILLIAMS, B. J., HANBY, A., MILLICAN-SLATER, R., NIJHAWAN, A., VERGHESE, E. & TREANOR, D. J. H. 2018a. Digital pathology for the primary diagnosis of breast histopathological specimens: an innovative validation and concordance study on digital pathology validation and training. 72, 662-671.
- WILLIAMS, B. J., LEE, J., OIEN, K. A. & TREANOR, D. 2018b. Digital pathology access and usage in the UK: results from a national survey on behalf of the National Cancer Research Institute's CM-Path initiative. *Journal of clinical pathology*, jclinpath-2017-204808.
- WILSON, E. C. J. P. 2015. A practical guide to value of information analysis. 33, 105-121.
- WOLFF, A. C., HAMMOND, M. E., SCHWARTZ, J. N., HAGERTY, K. L., ALLRED, D. C., COTE, R. J., DOWSETT, M., FITZGIBBONS, P. L., HANNA, W. M., LANGER, A., MCSHANE, L. M., PAIK, S., PEGRAM, M. D., PEREZ, E. A., PRESS, M. F., RHODES, A., STURGEON, C., TAUBE, S. E., TUBBS, R., VANCE, G. H., VAN DE VIJVER, M., WHEELER, T. M., HAYES, D. F. & AMERICAN SOCIETY OF CLINICAL ONCOLOGY/COLLEGE OF AMERICAN, P. 2007. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *Archives of Pathology & Laboratory Medicine*, 131, 18-43.
- WOLFF, A. C., HAMMOND, M. E. H., ALLISON, K. H., HARVEY, B. E., MANGU, P. B., BARTLETT, J. M., BILOUS, M., ELLIS, I. O., FITZGIBBONS, P. & HANNA, W. 2018. Human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline focused update. *Archives of pathology & laboratory medicine*.
- WOLFF, A. C., HAMMOND, M. E. H., HICKS, D. G., DOWSETT, M., MCSHANE, L. M., ALLISON, K. H., ALLRED, D. C., BARTLETT, J. M., BILOUS, M. & FITZGIBBONS, P. 2013. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *Journal of clinical oncology*, 31, 3997-4013.
- WOOLF, S. H., GROU, R., HUTCHINSON, A., ECCLES, M. & GRIMSHAW, J. J. B. 1999. Potential benefits, limitations, and harms of clinical guidelines. 318, 527-530.
- YAZIJ, H., TAYLOR, C. R., GOLDSTEIN, N. S., DABBS, D. J., HAMMOND, E. H., HEWLETT, B., FLOYD, A. D., BARRY, T. S., MARTIN, A. W., BADVE, S. J. A. I. & MORPHOLOGY, M. 2008. Consensus recommendations on estrogen receptor testing in breast cancer by immunohistochemistry. 16, 513-520.
- ZANOCCO, K., HELLER, M., ELARAJ, D. & STURGEON, C. 2013. Cost effectiveness of intraoperative pathology examination during diagnostic hemithyroidectomy for unilateral follicular thyroid neoplasms. *Journal of the American College of Surgeons*, 217, 702-710.
- ZARDAVAS, D., ADES, F. & DE AZAMBUJA, E. 2013. Clinical practice-changing trials: the HERA study paradigm. *Expert review of anticancer therapy*, 13, 1249-1256.

Appendices

Appendix A: Search strategies-systematic review of economic evaluations

Medline

1. exp Economics/
2. Economic*.mp.
3. Decision Tree.mp.
4. Markov*.mp.
5. Model*.mp.
6. Simulat*.mp.
7. 5 and 6
8. Cost.mp.
9. *"costs and cost analysis"/ or *"cost allocation"/ or *cost-benefit analysis/ or *"cost control"/ or *"cost savings"/ or *"cost of illness"/ or *health care costs/ or *direct service costs/ or *drug costs/ or *employer health costs/ or *hospital costs/ or *health expenditures/
10. *Decision Trees/
11. *models, statistical/ or *models, economic/ or *models, econometric/
12. 1 or 2 or 3 or 4 or 7 or 8 or 9 or 10 or 11
13. Tumor.mp.
14. Tumour.mp.
15. *Neoplasms/
16. Malignan*.mp.
17. *Carcinoma/
18. Cancer*.mp.
19. *"Early Detection of Cancer"/
20. 13 or 14 or 15 or 16 or 17 or 18 or 19
21. *pathology/ or *pathology, clinical/ or *pathology, molecular/ or *pathology, surgical/ or *telepathology/
22. Pathologist.mp.
23. Histopathology.mp.
24. Digital Pathology.mp.
25. Telepathology.mp.
26. 21 or 22 or 23 or 24 or 25
27. 12 and 20 and 26
28. limit 27 to english language ;l
29. limit 28 to yr="2000 -Current"

Embase

1. exp economics/
2. exp health economics/
3. Economic*.mp.
4. exp economic evaluation/
5. exp "cost utility analysis"/ or exp "cost benefit analysis"/ or exp "cost minimization analysis"/ or exp "cost effectiveness analysis"/
6. exp "health care cost"/
7. Cost*.mp.
8. *model/
9. exp statistical model/

10. *"decision tree"/
11. Decision Tree.mp.
12. Markov.mp.
13. *neoplasm/ or *metastasis/ or *"precancer and cancer-in-situ"/
14. Malignan*.mp.
15. Tumour.mp.
16. *carcinoma/
17. "Cancer*".m_titl.
18. "Tumor*".m_titl.
19. *pathology/
20. *pathologist/
21. *histopathology/
22. *telepathology/
23. Digital Pathology.mp.
24. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12
25. 13 or 14 or 15 or 16 or 17 or 18
26. 19 or 20 or 21 or 22 or 23
27. 24 and 25 and 26
28. limit 27 to english language
29. limit 28 to yr="2000 -Current"

Econlit

This strategy did not include search terms related to economic evaluations as this is a subject specific database.

	Query
S11	S6 AND S9
S10	S6 AND S9
S9	S7 OR S8
S8	pathologist
S7	pathology
S6	S1 OR S2 OR S3 OR S4 OR S5
S5	oncology
S4	Malignan*
S3	tumours or cancer or neoplasms
S2	Carcinoma*
S1	Cancer*

Web of Science

7 #5 AND #4 AND #3
 Refined by: LANGUAGES: (ENGLISH)
 Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years
 # 6 #5 AND #4 AND #3
 Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years
 # 5 TS=(Tumour* OR Tumor* OR Cancer* OR Malignan* OR Carcinoma OR Neoplasm)
 Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years
 # 4 TS=*Patholog*
 Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years
 # 3 #2 OR #1
 Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years

2 TI=Cost*
 Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years
 # 1 TI=Economic*
 Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years

CRD Database

NHSEED

```

1      MeSH DESCRIPTOR Early Detection of Cancer EXPLODE ALL TREES
2      (Cancer*) IN NHSEED
3      (Tumor*) IN NHSEED
4      (Tumour*) IN NHSEED
5      (Neoplasm*) IN NHSEED
6      (Carcinoma*) IN NHSEED
7      (Malignan*) IN NHSEED
8      #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7
9      (Patholog*) IN NHSEED
10     (Histopatholog*) IN NHSEED
11     MeSH DESCRIPTOR Pathology EXPLODE ALL TREES
12     #9 OR #10 OR #11
13     #8 AND #12
  
```

HTA

Line	Search
1	MeSH DESCRIPTOR Health Care Economics and Organizations EXPLODE ALL TREES
2	MeSH DESCRIPTOR Models, Economic EXPLODE ALL TREES
3	MeSH DESCRIPTOR Value of Life EXPLODE ALL TREES
4	(Cost Utility) IN HTA
5	(Cost Benefit) IN HTA
6	(Cost Effectiveness) IN HTA
7	MeSH DESCRIPTOR Decision Trees EXPLODE ALL TREES
8	(Decision Tree) IN HTA
9	(Markov*) IN HTA
10	(Economic Evaluation) IN HTA
11	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10
12	MeSH DESCRIPTOR Early Detection of Cancer EXPLODE ALL TREES
13	(Cancer*) IN HTA
14	(Tumor*) IN HTA
15	(Tumour*) IN HTA
16	(Neoplasm*) IN HTA
17	(Carcinoma*) IN HTA
18	(Malignan*) IN HTA
19	#12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18
20	MeSH DESCRIPTOR Pathology EXPLODE ALL TREES
21	(Patholog*) IN HTA
22	(Histopatholog*) IN HTA
23	#20 OR #21 OR #22
24	#11 AND #19 AND #23


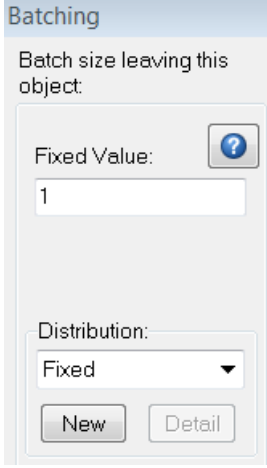




DARE

Line	Search
1	MeSH DESCRIPTOR Health Care Economics and Organizations EXPLODE ALL TREES
2	MeSH DESCRIPTOR Value of Life EXPLODE ALL TREES
3	MeSH DESCRIPTOR Models, Economic EXPLODE ALL TREES
4	(Cost Utility) IN DARE
5	(Cost Benefit) IN DARE
6	(Cost Effectiveness) IN DARE
7	MeSH DESCRIPTOR Decision Trees EXPLODE ALL TREES
8	(Decision Tree) IN DARE
9	(Markov*) IN DARE
10	(Economic Evaluation) IN DARE
11	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10
12	MeSH DESCRIPTOR Early Detection of Cancer EXPLODE ALL TREES
13	(Cancer*) IN DARE
14	(Tumor*) IN DARE
15	(Tumour*) IN DARE
16	(Neoplasm*) IN DARE
17	(Carcinoma*) IN DARE
18	(Malignan*) IN DARE
19	#12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18
20	MeSH DESCRIPTOR Pathology EXPLODE ALL TREES
21	(Patholog*) IN DARE
22	(Histopatholog*) IN DARE
23	#20 OR #21 OR #22
24	#11 AND #19 AND #23

Appendix B: Data extraction form - systematic review of economic evaluations

Name:		For completion by the reviewer	
Notes			
Bibliographic details	Author		
	Year		
	Journal		
	Publication Type		
	Additional Comments		
General details	Country		
	Study site(s)		
	Additional comments		
Target population	Cancer type		
	Sample size		
	Additional comments		
Pathology	Intervention details		
	Comparator details		
	Additional comments		
Economic evaluation	Perspective(s)?		
	Approach		
	Design		
	Time horizon		
	Additional comments		
	Outcome(s) measured	<i>Methods</i>	<i>Sources</i>
	Total intervention outcome		
	Total comparator outcome		
	Additional Comments		
	Unit cost(s)	<i>Methods</i>	<i>Sources</i>
	Resource Use	<i>Methods</i>	<i>Sources</i>
	Currency		
	Financial year		
	Discounting		
	Total intervention costs		
	Total comparator costs		
	Additional Comments		
	Uncertainty analysis		
	Decision rule		
	Cost-effectiveness results		
Additional comments			

Appendix C: Simul8 building blocks

Building Block	Description	Image
Activity	An entity has something done to it here	
Batching	Method in which we can split one entity so that it moves through 2 or more activities simultaneously e.g. a case has both ER and HER2 testing simultaneously	
End	Where the entity leaves the model	
Entity	The work item in the system, here these are suspected breast cancer cases	
Labels	These are used to attach attributes such as test results to an entity which are used to control how it moves through the model	NA
Queues	An entity waits here to be processed	
Routing	These rules tell simul8 where the entity should go	NA
Start	Where an entity enters the model	
Visual logic	Is the Simul8 language allowing detailed codes to be built into the model to control how it behaves	NA

Appendix D: List of search terms

Key words listed in the five published papers

-Breast Cancer	-Breast Neoplasms
-HER-2	-Fluorescence in Situ Hybridization
-Herceptest	-HER2
-Reproducibility	-Targeted Therapy
-Subjective Analysis	-Image Analysis
-Digital Image Analysis	-IHC
-FISH	-Virtual Slides
-ACIS	-Automated IHC Method
-Immunohistochemistry	-HER2 FISH test
-Automation	-Primary Invasive Breast Cancer
-Trastuzumab	-Survival
-Herceptin	-Visual IHC Method

Key words identified through screening the titles and abstracts of the five published papers

-Immunohistochemical Staining	-Comparison	-Misinterpreted
-Concordance/Concordant	-Visual	-Standardise
-HER-2 targeted therapy	-Assessment	-Interpretation
-Manual	-Impact	-Diagnostic tool
-Inter observer reproducibility	-Outcome	-Determination
-Accurate	-Computerized Image Analysis	-Variability
-Automated	-Agreement	-Uncertainty
-Imaging system	-Visual Scoring	-Reliability
-Correlation/Correlated	-Algorithm	-Consistent
-Performance	-Evaluation	-Classify
-Positive predictive value	-Results	-Computer Aided
-Quality	-Biomarker	-Quantitative
-Subjective	-Conventional Testing	-Reference
-Expression	-HER2 protein expression	
-Scoring/scored/score	-Semi-quantitative	
-Inter observer	-Digital Images	
-Discrepancies	-Evaluation	
-Comparable/Compare	-Correlation	
-Performance	-Kappa	
-Consensus	-Visual Evaluation	

Appendix E: Search terms used in initial phase search strategy

	Keyword	Related Terms	Truncations
Population	Breast	Breast Cancer Breast Neoplasms Primary Invasive Breast Cancer	
	HER2	HER-2 Biomarker HER2 protein expression	
	IHC	Herceptest Immunohistochemistry Immunohistochemical staining	
Intervention	IHC DP IA	Digital Image Analysis Imaging system Computerised Digital Images Virtual Slides Algorithm Computer Automation/Automated Aided Automated IHC method Quantitative	Imag* Automat* Computer*
Comparator	IHC CM	Subjective Analysis Visual IHC method Visual Manual Subjective Conventional Testing Semi-quantitative	
Reference Standard	FISH	Fluorescence in situ hybridisation Reference	
Outcome	Accuracy	Kappa Consistent Accurate Reliable Uncertainty Performance Variability Agreement Sensitivity Specificity Discrepancies Comparable/Compare/Comparison Reproducibility Misinterpreted Correlation/Correlated Positive predictive value Scoring/scored/score Inter observer	Concordan* Correlat* Scor* Compar* Accura*
	Herceptin	Trastuzumab Targeted Therapy HER-2 targeted therapy	

Appendix F: Search strategies - systematic review of HER2 accuracy

Medline

- 1 exp Breast Neoplasms/
- 2 Breast Neoplasm.mp.
- 3 Breast Cancer.mp.
- 4 1 or 2 or 3
- 5 exp Receptor, ErbB-2/
- 6 exp Receptor, Epidermal Growth Factor/
- 7 HER2.mp.
- 8 HER-2.mp.
- 9 5 or 6 or 7 or 8
- 10 exp Immunohistochemistry/
- 11 IHC.mp.
- 12 Immunohistochemi*.mp.
- 13 10 or 11 or 12
- 14 exp Image Processing, Computer-Assisted/
- 15 exp Image Interpretation, Computer-Assisted/
- 16 exp Algorithms/
- 17 Digital.mp.
- 18 exp Automation/
- 19 Automat*.mp.
- 20 Virtual.mp.
- 21 Algorithm.mp.
- 22 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21
- 23 exp In Situ Hybridization, Fluorescence/
- 24 Fluorescence in situ hybridisation.mp.
- 25 FISH.mp.
- 26 23 or 24 or 25
- 27 4 and 9 and 13 and 22 and 26

Embase

- 1 exp breast tumor/
- 2 Breast Cancer.mp.
- 3 Breast Neoplasm.mp.
- 4 1 or 2 or 3
- 5 HER2.mp.
- 6 *epidermal growth factor receptor 2/
- 7 HER-2.mp.
- 8 5 or 6 or 7
- 9 exp immunohistochemistry/
- 10 IHC.mp.
- 11 immunohistochemistry.mp.
- 12 9 or 10 or 11
- 13 *computer assisted diagnosis/
- 14 *computer analysis/
- 15 exp laboratory automation/
- 16 exp digital imaging/
- 17 exp digital microscope/
- 18 exp digital slide scanner/
- 19 image processing/

20 exp image analysis/
 21 exp imaging system/
 22 imaging system.mp.
 23 exp microscope image/
 24 exp quantitative diagnosis/
 25 exp quantitative histochemistry/
 26 virtual slides.mp.
 27 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26
 28 FISH.mp.
 29 exp fluorescence in situ hybridization/
 30 Fluorescence in situ hybridisation.mp.
 31 28 or 29 or 30
 32 4 and 8 and 12 and 27 and 31

Web of Science

20 #19 AND #16 AND #13 AND #7 AND #4
 Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years
 # 19 #18 AND #17
 Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years
 # 18 TOPIC: (Cancer) OR TOPIC: (Neoplasm)
 Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years
 # 17 TOPIC: (Breast*)
 Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years
 # 16 #15 OR #14
 Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years
 # 15 TOPIC: (FISH)
 Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years
 # 14 TOPIC: (Fluorescence in situ hybridisation)
 Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years
 # 13 #12 OR #11 OR #10 OR #9 OR #8
 Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years
 # 12 TS=(automation)
 Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years
 # 11 TS=(quantitative diagnosis)
 Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years
 # 10 TS=(virtual slide)
 Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years
 # 9 TS=(Image analysis)
 Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years
 # 8 TS=(Digital Pathology)
 Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years
 # 7 #6 OR #5
 Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years
 # 6 TOPIC: (IHC)
 Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years
 # 5 TS=(immunohistochemistry)
 Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years
 # 4 #3 OR #2 OR #1
 Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years
 # 3 TOPIC: (Epidermal growth factor receptor 2)
 Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years

2 TOPIC: (HER-2)
 Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years
 # 1 TOPIC: (HER2)
 Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years

PubMed

#10 ((((((FISH) OR Fluorescence in situ hybridisation)) AND (((Digital Pathology) OR Image Analysis) OR automat*) OR Virtual Slides)) AND ((Immunohistochemi*) OR IHC)) AND (((HER2) OR HER-2) OR Human Epidermal Growth Factor Receptor 2)) AND ((Breast) AND ((Neoplasm) OR Cancer))
 #9 (FISH) OR Fluorescence in situ hybridisation
 #8 (((Digital Pathology) OR Image Analysis) OR automat*) OR Virtual Slides
 #7 (Immunohistochemi*) OR IHC
 #6 ((HER2) OR HER-2) OR Human Epidermal Growth Factor Receptor 2
 #5 (Breast) AND ((Neoplasm) OR Cancer)
 #4 (Neoplasm) OR Cancer
 #3 Breast
 #2 Neoplasm
 #1 Cancer

Cochrane Library

#1 Breast Cancer
 #2 MeSH descriptor: [Breast Neoplasms] this term only
 #3 #1 or #2
 #4 Her-2
 #5 HER2
 #6 Human epidermal growth factor receptor 2
 #7 MeSH descriptor: [Genes, erbB-2] this term only
 #8 #4 or #5 or #6 or #7
 #9 IHC
 #10 immunohistochem*
 #11 #9 or #10
 #12 FISH
 #13 Fluorescence in situ hybridisation
 #14 MeSH descriptor: [In Situ Hybridization, Fluorescence] explode all trees
 #15 #12 or #13 or #14
 #16 Imag*
 #17 Digital
 #18 MeSH descriptor: [Image Interpretation, Computer-Assisted] explode all trees
 8722
 #19 algorithm
 #20 automat*
 #21 MeSH descriptor: [Algorithms] this term only
 #22 MeSH descriptor: [Automation, Laboratory] explode all trees
 #23 quantitative
 #24 MeSH descriptor: [Diagnosis, Computer-Assisted] explode all trees
 #25 #16 or #17 or #18 or #19 or #20 or #21 or #22 or #23 or #24
 #26 #3 and #8 and #11 and #15 and #25

Appendix G: Data extraction form - systematic review of HER2 accuracy

	For completion by the reviewer		
Notes			
Name of reviewer			
Bibliographic Details	Title		
	Author		
	Year		
	Journal		
	Publication Type		
	Additional Comments		
Sample Size			
IHC CM	Assay		
	Threshold	0	
		1+	
		2+	
		3+	
	Additional Comments		
IHC DP	Scanner		
	Algorithm		
	Measurement		
	Additional Comments		

FISH	Assay							
	Threshold	-ve						
		Equi						
		+ve						
	Additional Comments							
2x2 Tables	Fish	IHC CM			IHC DP			
		-ve	Equi	+ve	-ve	Equi	+ve	
	-ve							
	Equi							
	+ve							
<i>Additional Comments</i>								
Outcomes	<i>IHC CM</i>							
	Sensitivity							
	Specificity							
	%2+							
	Additional Comments							
	<i>IHC DP</i>							
	Sensitivity							
	Specificity							
	%2+							
	Additional Comments							